UNIVERSITY OF CALIFORNIA

Santa Barbara

Genetic analysis of developmental fidelity in *C. elegans*

A Dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in Molecular, Cellular, and Developmental Biology

by

Melissa R. Alcorn

Committee in charge:

Professor Joel Rothman, chair

Professor Kathleen Foltz

Professor William Smith

Professor Thomas Turner

March 2020

The dissertation of Melissa R. Alcorn is approved.

_____

Kathleen Foltz

_____

William Smith

_____

Thomas Turner

_____

Joel Rothman, Committee Chair

March 2020

Genetic analysis of developmental fidelity in *C. elegans*

ACKNOWLEDGEMENTS

I would like to dedicate this work to my grandmother, Myrna Eddy. She was my biggest advocate and I miss her terribly. More than anything, she wanted me to be independent – which meant getting an education. The only remarkable thing about this thesis is the sheer amount of data and work (from multiple people) it reflects, and even though this work is not a complete story I hope that she would be proud of me. I would also like to thank my advisor Dr. Joel Rothman and all members of the Rothman lab for their patience, guidance and support.

And of course, my biggest supporter – Smokey Joe, a blue-ribbon kitty.

Vita of Melissa R. Alcorn
March 2020

EDUCATION

Doctor of Philosophy in Molecular, Cellular and Developmental Biology, University of California – Santa Barbara, Santa Barbara, CA March 2020

Bachelor of Science in Chemistry, Loyola University, Chicago, June 2012

SPECIALIZED TRAINING AND RESEARCH EXPERIENCE

Short Course in Systems Biology: Morphogenesis and Spatial Dynamics, University of California, Irvine, Irvine, CA. Course coordinators: Felix Grün, PhD, Arthur D. Lander, M.D., Ph.D. (2016)
Research assistant in Morse Lab, University of California, Santa Barbara, Santa Barbara, CA. PI Daniel Morse, Ph.D. (2011)
Research assistant, M.R.S.A. Research Center, University of Chicago Medical Center, Chicago, IL; PI Robert Daum, M.D. (2010)

PUBLICATIONS AND CONTRIBUTIONS

Torres Cleuren, Y. N., Ewe, C.K., Chipman, K.C., Mears, E.R., Wood, C.G., Al-Alami, C.E., **Alcorn, M.R.,** Turner, T.L., Joshi, P.M., Snell, R.G., Rothman, J.H. (2019) Extensive intraspecies cryptic variation in an ancient embryonic gene regulatory network. *Elife* 8. doi:10.7554/elife.48220.

Alcorn, M. R., Callander, D. C., López-Santos, A., Torres Cleuren, Y. N., Birsoy, B., Joshi, P. M., Santure, A. W. & Rothman, J. H. (2016) Heterotaxy in *Caenorhabditis*: widespread natural variation in left–right arrangement of the major organs. *Philos. Trans. R. Soc. B Biol. Sci.* **371**. (doi:10.1098/rstb.2015.0404)

Callander, D. C., **Alcorn, M. R.,** Birsoy, B. & Rothman, J. H. (2014) Natural reversal of left-right gut/gonad asymmetry in *C. elegans* males is independent of embryonic chirality. *Genesis* **52**, 581– 587. (doi:10.1002/dvg.22762)

SCIENTIFIC PRESENTATIONS

Melissa Alcorn, Yamila N Torres Cleuren, Matt Cieslak, Joel Rothman. Applying machine learning tools and natural variation to identify loci underlying genetically complex traits: comparison with GWAS and QTL mapping. Presented a poster at International Worm Meeting, University California, Los Angeles, Los Angeles, CA, June 25, 2017.

Melissa Alcorn, Davon Callander, Bilge Birsoy, Joel Rothman. The propensity for stochastic programmed cell death during male tail development varies across *C. elegans* isotypes. Presented a poster at International Worm Meeting, University California, Los Angeles, Los Angeles, CA, June 26, 2017.

Melissa Alcorn, Davon Callander, Bilge Birsoy, Matthew Cieslak, Joel Rothman. The propensity for stochastic programmed cell death during male tail development varies across *C. elegans* isotypes. Poster presentation at Regional Systems Biology Conference, University California, Irvine, Irvine, CA, January 30, 2016.

Melissa Alcorn, Davon Callander, Bilge Birsoy, Matthew Cieslak, Agustín López Santos, Joel Rothman. Developmental fidelity varies in *C. elegans* males. Presented a short talk at International Worm Meeting, University California, Los Angeles, Los Angeles, CA, June 26, 2015.

Melissa Alcorn, Davon Callander, Bilge Birsoy, Matthew Cieslak, Agustín López Santos, Joel Rothman. The propensity for stochastic PCD during male tail development varies across *C. elegans* isotypes**.** Presented poster at International Worm Meeting, University California, Los Angeles, Los Angeles, CA, June 26, 2015.

AWARDS AND FELLOWSHIPS

2015-16	Special Fellowship in the STEM Disciplines ($24,000) Doreen J. Putrah Cancer Research Foundation Conference Fellowship Grant ($1000) UCSB Graduate Division Association Travel Grant ($200) Travel/Course Fees Fellowship for National Short Course in Systems Biology at University of California, Irvine ($1000)
2014	MCDB continuing Fellowships (tuition and support)
2013	MCDB continuing Fellowships (tuition and support)

TEACHING EXPERIENCE

2012-2019	Teaching Assistant at University of California, Santa Barbara. Topics taught include: introductory biology, developmental biology, systems biology (coding in Matlab), biochemistry, biochemistry techniques (SDS-PAGE, Western blotting, protein purification assays), molecular genetics laboratory techniques (RNAi, PCR, light and fluorescence microscopy), experimental design and critical thinking

ABSTRACT

Genetic analysis of developmental fidelity in *C. elegans*

By Melissa R. Alcorn

Organismal development relies on precise temporal and spatial regulation of numerous processes to ensure proper formation of the body plan including the enforcement of symmetric and asymmetric systems. Gene regulatory networks (GRNs) are organized in such a way as to ensure fidelity of each process, or to ensure a precise and accurate phenotypic outcome despite the noise inherent in biological systems. This leads to nontrivial questions, including what genetic mechanisms, if any, are in place to ensure fidelity and how is a GRN as a whole affected by genetic variation in its individual components. Further, how does such variation lead to phenotypic variation, especially with complex or multigenic traits that rely on multiple inputs. Here I describe the application of quantitative genetic methods to identify loci associated with fidelity of two separate but important developmental processes in the nematode *C. elegans*: 1) Left-right (L/R) asymmetry of the gut-gonad orientation, and 2) proper implementation of programmed cell death (PCD) events that typically sculpt bilaterally symmetric sensory rays of the male tail, which are necessary for copulation. We found significant phenotypic variation for both phenotypes in males from *C. elegans* isotypes and closely related gonochoristic and hermaphroditic species, suggesting that variation in these traits is not solely attributable to lab domestication or differences in selective pressure(s) on males (Chapters 3 and 4). For association

mapping, we use two sources of genetic and phenotypic variation – a globally diverse collection of *C. elegans* isotypes and recombinant inbred lines (RILs) generated from isotypes with either consistently high or low phenotypic variance for additional traits. We apply genome wide association studies (using linear mixed model design for highly inbred model organisms) and quantitative trait locus (QTL) mapping to identify causal genetic regions. We report that the missing rays or ray defects originally noted by Sulston were nearly suppressed in canonical programmed cell death mutants in the reference strain in N2 (Chapter 2); this was also the case in some *C. elegans* isotypes with a high propensity for defects when the endogenous pro-apoptotic factor EGL-1, a BH3-only domain protein, was knocked down using RNAi (Chapter 3), suggesting there is standing genetic variation in mis-regulation of the PCD pathway. GWAS and linkage disequilibrium analysis of significant SNPs suggest a ~3 Mb region on chromosome II is associated with stochastic cell death in 87 unique *C. elegans* isotypes. QTL mapping with RILs made from isolates QX1211 and AB4 further identified a much larger region on chromosome II, in addition to QTL on chromosomes III and X, that explains ~ 55% of natural variation in this trait (Chapter 3). In the case of the observed heterotaxy – the improper arrangement of at least one but not all visceral organs, GWAS identified two QTL on chromosomes II and III that affect only males, suggesting that different QTLs and other genetic/cellular mechanisms mediate natural variation in males compared to those affecting hermaphrodite heterotaxy (Chapter 4). 95% confidence intervals and linkage disequilibrium analysis indicate that these QTL for both traits are relatively large (compared to the length of the chromosome (Chapters 3 and 4)), and near-isogenic lines (NILs) excluded a small

~1.5 mb region on chromosome II. We applied variant-effect prediction (VEP) analysis of the remaining QTL on chromosomes II, III and X as an alternative to identified possible candidates (Chapter 3).

In parallel to the standard approaches, we have also applied a machine learning pipeline using ElasticNet Regression as a complementary approach for identifying genetic loci that may be missed as a result of small effect size, population structure, or low statistical power for naturally varying traits in *C. elegans* isotypes (Chapter 5). Our approach was similar to its use in multi-parental inbred mice lines but was combined with, and enhanced by, the genetic toolkit of *C. elegans*. For example, *C. elegans* isotypes can be used to screen a large amount of candidate genes for functional studies in polygenic backgrounds. We used the previously described, naturally varying requirement for SKN-1 – a maternally-loaded transcription factor necessary for endoderm specification - to test the performance of ElasticNet regression in identifying biologically confirmed regions from other standard approaches like QTL mapping and GWAS. ElasticNet regression was able to identify all loci identified by GWAS and QTL mapping with a set of recombinant inbred lines made with isotypes N2 and MY16 (Spearman's rho, $R^2 = 0.55$, $p = 6.5 \times 10^{-5}$), in addition to novel loci on chromosomes V and X. While this model is significant statistically, isogenic line testing will be necessary to validate this method as a valid alternative to more traditional association mapping methods (Chapter 5).

Table of Contents

LIST OF FIGURES AND TABLES

Chapter One

Introduction: gene regulatory networks and robustness of development in *C. elegans*

Canalization by gene regulatory networks (GRNs) ensures developmental fidelity

Successful development is a nontrivial task and viability of a species requires that specialized cells are precisely and robustly made in each individual, despite a constant influx of environmental stressors or millions of years' worth of accumulated cryptic genetic variants. Fidelity relies on precise spatiotemporal gene expression, which is dependent on organization of gene regulatory networks (GRNs) for concerted development. In Waddington's model, a zygote has an intended developmental path on a highly contoured landscape where the contours of the landscape are determined by the intersection of gene regulatory networks underneath (1, 2). Trait canalization also implies that there is some sort of active process that keeps the zygote on its trajectory for a specific phenotypic outcome. While processes like DNA replication, aminoacyl-tRNA charging, and repair of genome damage are considered high fidelity and have known error-correcting mechanisms (3–5), not many genes, GRNs or mechanisms are known that may regulate or ensure precise spatiotemporal gene expression for successful development. While there is a compelling case to be made for the existence of an 'error-correcting' or developmental noise buffering mechanism, there are not many examples to date. One of the most well characterized mechanisms involves the chaperone protein Hsp90, where blocking its activity in different mutant backgrounds causes a different, variable phenotype - suggesting it is capable of buffering noise that may occur from accumulation of random cryptic genetic variants for a wide variety of developmental processes (6–8). An example of a canalized trait in

nematodes is vulval development – where a combination of Ras, Notch and EGF signaling amongst vulval precursor cells (VPCs PnP.3-8) produces a highly stereotyped vulva during the early L4 stage and cell-fate pattern is consistent in closely related species (reviewed in (9)). When grown on plates at standard conditions (20˚C), VPCs Pn3.P-8 adopt a highly reproducible pattern of cell-fates referred to as "3˚-3˚-2˚-1˚-2˚-3˚", where Pn6.P adopts primary (1˚) cell-fate and becomes the inner vulval cell, Pn5.P and 7 adopt secondary (2˚) fate and become outer-vulval cells and PnP.3,4 and 8 become tertiary (3˚) non-vulval cells. Perturbation of the system, through genetic mutations, physical laser ablation of cells and signaling centers, or by growing under different conditions reveal compensatory mechanisms that are in play at various levels. For example, Wnt signaling is the predominant signaling pathway when N2 worms are cultured in liquid as opposed to growth on plates where EGF-RAS pathway is the primary signaling pathway, while the pattern of VPC fates remains the same. Laser ablation of the anchor cell, the primary source of EGF signaling, reveals different compensatory patterns of cell-fates in the remaining VPCs in closely related species. When LIN-3/EGF signaling is removed by ablation of the anchor cell In N2, the remaining VPCs adopt a '2˚ - 3˚ - 2 'pattern in cells Pn5.5-7, while in the hermaphroditic species *C. briggsae* – when cells adopt a '2˚ 2˚ 2˚', suggesting that there is and is a variable requirement for EGF signaling in the GRN regulating vulval cell-fate induction, and cryptic genetic variants affect the gene regulatory networks involved as this variation is only revealed after perturbation of the system (10, 11).

Mis-regulation of GRNs can lead to pathological conditions

Successful development is a nontrivial task and viability of a species requires that specialized cells are precisely and robustly made in each individual. Successful development requires precise deployment of specific GRNs to consistently produce a phenotype with low variation despite perturbations like changes in temperature, nutrient availability, or millions of years of accumulated heritable changes in DNA (12, 13, 13, 14). Errors in development can lead to major structural differences in organ function and have been associated with pathological conditions in humans. Structural changes include apparently benign medical conditions like *situs inversus* where all major internal organs are rearranged in a perfect mirror image of what is considered 'dextral' or wild type (such as the lungs, kidneys and heart (15, 16)) or more troublesome conditions where one or more (but not all) major organs are misaligned – a condition referred to as heterotaxy and is often associated with ciliopathies (15). Additionally, there is significant phenotypic variation seen in the brain/central nervous system of patients with neurodegenerative diseases and neurodevelopmental disorders. One GRN that is highly associated neurodegenerative disease and neurodevelopmental disorders is the Programmed Cell Death (PCD) pathway, which when mis-regulated during development or later in life can have major structural effects on the nervous system (17). Some neuropsychiatric conditions like schizophrenia (SCZ), bipolar disorder (BP), and autism, have been linked to increased cellular stress and rates of apoptosis that change the local structure in several brain regions that mediate speech, perception and other cognitive abilities (18–21). fMRI evidence suggests there are major

structural changes in the brain in patients with Schizophrenia, such as overall loss of cortical volume when compared to case-control patients (22). The pervasive and debilitating effects of these conditions are a major reason for the significant effort to study how GRNs are coordinated over developmental time periods and assemble a human connectome in the hopes that we may better understand the relationship between gene expression and the underlying structural differences in GRNs (23).

Structure and function of GRNs

GRNs are a collection of genetic interactions typically associated with a particular ligand, and interactions with different GRN at specific developmental timepoints are also associated with phenotypic outcomes (24, 25). They encode the mechanisms that form the body-plan for a species and wiring diagrams are used to depict the interactions between the individual genes that form gene regulatory networks (26). GRNs have a hierarchical organization where morphogenesis events are initiated typically by transcription factors binding to cis-regulatory regions of other nodes (genes) in the network and these transcriptional regulatory relationships in GRNs have been called “kernels” (27). The output of these key transcriptional regulators and kernels ultimately activate “gene differentiation batteries,” comprised of protein-coding genes that underlie the final stages of morphogenesis and differentiation (25). Transcription factors can be coopted by novel developmental processes (or vice versa) and their role can vary significantly even in closely related species, like the requirement for SKN-1 and POP-1 transcription factors in in endoderm specification between *C. elegans* and *C. briggsae* or the variation in requirement for

Delta/Notch signaling in endomesoderm specification of two closely related sea urchin species (28, 29). It is of interest to depict how GRN are wired together as activators and repressors for a particular phenotype, their epistatic relationships and how GRN may respond to particular perturbations such as environmental stress or removal of signaling centers (27, 30).

Methods to identify and describe GRN architecture

Quantitative genetics is one way to study the relative importance of individual nodes in a GRN that contribute to a particular phenotype and to describe the larger phenotypic space. It also has potential biomedical relevance: accurately predicting the phenotypic effect of a complex set of mutations or genetic variants is challenging, and an understanding of such predictions can be refined by studying quantitative variation in well-described networks of model organisms. Genome-wide association studies (GWAS) and quantitative trait loci (QTL) mapping are standard methods to identify regions of the genome associated with a particular phenotype (31, 32). The goal is to identify any and all loci associated with a phenotype *in silico* and then confirm causal sequence variants *in vivo*. As computational power increases and whole-genome sequencing becomes less expensive, large amounts of data can be obtained for association mapping (33). Single nucleotide polymorphisms (SNPs) can typically be associated with a particular phenotype through linear mixed-model equations with kinship matrices (34–39). From many such studies, it has become apparent that the majority of quantitative traits are multi-genic, and effects of single variants can often

be hidden by epistatic interactions. Thus, adapting quantitative methods to address this issue is an active field of investigation of importance to a wide range of biomedical and fundamental questions in developmental biology (See Chapter 5).

C. elegans as a model system for quantitative genetics and fidelity of developmental processes

Naturally varying traits in any system are generally dependent on interactions with multiple loci that may not always contribute equally to the variation of a particular trait. Traditional forward genetics methods, such as genome-wide functional RNAi screening for single mutations or isolation of near-isogenic/homozygous lines from mutagenesis, can fail to reveal such complex multi-genic interactions such as the role of Wnt signaling in the specification of the vulva precursor cell Pn3.P or the role of *pha-1* and *sup-35* in development as a genetic selfish element (1–3). The alternative of quantitative genetics approaches allows effect-size of multiple loci to be calculated. These methods can also describe how genetic variation shapes the phenotypic space of a particular trait. The natural variation in two highly heritable *C. elegans* traits that we report here can be used as models for understanding complex, polygenic diseases in a simpler, genetically tractable organism. Recently, a growing collection of globally diverse *C. elegans* isotypes has been sequenced and curated (40). The rich genetic diversity and phenotypic diversity, combined with an already expansive genetic tool kit, make *C. elegans* a good model for quantitative genetics. Such studies have identified causal variants for a number of naturally quantitatively varying traits (41–46). The rapid and stereotyped development, ease of imaging of individual cells by

Differential Interphase Contrast (DIC) microscopy, and robust development in large isogenic populations are further assets to studying developmental variation. One of the salient features of *C. elegans* as a model organism is its highly determinate lineage that is tightly genetically controlled, which has allowed for detailed analysis of its development and connectome at single-cell resolution (47–50). Taken together, these characteristics make *C. elegans* a highly attractive organism for precise quantification of variation in nearly any measurable phenotype. The natural variation in two highly heritable *C. elegans* traits that we report here – propensity for heterotaxy and stochastic sensory cell death in mail tails, can be used as models for understanding complex, polygenic diseases in a simpler, genetically tractable organism.

Male tail development in *C. elegans*

The male tail is a specialized mating structure comprised of a modified rectal epithelial chamber called the proctodeum, a pair of sclerotic spicules, and nine bilateral pairs of sensory rays embedded in a cuticular fan (Figure 1.1.). Bilateral pairs of sensory rays have unique morphology, neuropeptide expression and are important for relaying sensory information from the environment required for different aspects of successful mating behavior (51, 52). All rays are derived post-embryonically from lateral seam cells V5 (L/R), V6 (L/R) and T (L/R). These cells rapidly divide near the mid-early L3 stage where in addition to the sensory rays, the descendants become the sex-specific muscles, interneurons, neurons and epithelial sensory structures necessary for successful copulation (53). During the late L4

stage, 18 independent ray precursor cells (Rn) follow a highly stereotyped sub-lineage to produce: two sensory or motor-sensory neurons (RnA or RnB), a structural support cell (Rnst) and a hypodermal cell (Rn.p). Within this shared sub-lineage, the sister-cell of the RnA neuron is fated to die (Figure 1.1) (49, 52, 54, 55). The structural support cell (Rnst) forms a tube around the dendritic endings of the two neurons, which are then wrapped by the hypodermal cell. Hypodermal cells (Rn.p) from rays 1-5 (R1.p-5) then fuse to form the male posterior seam cell (SET cell) and those from rays 7-9 fuse with hyp7. Both hyp7 and the SET cell are encased by the cuticle (54).

The evolutionarily conserved cell death GRN in *C. elegans*

In mammalian systems, cells that undergo canonical apoptosis follow a stereotyped, organized dismantling of the cell to prevent inflammation or damage to neighboring cells. In the intrinsic cell death pathway, cellular stress (such as damaged mitochondria via oxidative stress) triggers conformational changes in Bcl-2-family proteins. These active proteins can then permeabilize the outer mitochondrial membrane, releasing cytochrome C into the cytosol. If damage is significant, cytochrome C reaches a critical threshold at which it activates the formation of the apoptosome: pro-caspase-9 is converted to activated CASPASE-9, which can then proteolytically activate executioner caspases that cleave multiple cellular substrates, promote nuclear DNA degradation, mitochondrial fission, inhibition of survival/growth signals and engulfment (56, 57). While CASPASES-3,6 and 7 are considered executioner caspases that trigger DNA fragmentation in apoptosis, CASPASE-3 also has significant non-apoptotic roles such as axonal pruning (58).

C. elegans is a fundamental model organism to the study of PCD: cell corpses have a distinct morphology that can be readily identified using differential interference contrast (DIC) microscopy. Genetically they have evolutionarily conserved core apoptotic machinery: EGL-1 – a pro-apoptotic BH3-only domain protein interrupts anti-apoptotic interaction between CED-9 (a Bcl-2 homolog) and Apoptotic protease activating factor-1 (Apaf-1) ortholog CED-4 and caspase CED-3, an ortholog to human CASP2 (caspase-2) (59, 60). The majority of developmental PCD in *C. elegans* is often thought of as the combination of three distinct phases: specification, execution and engulfment. In cells specified to die, unique combination of transcription factors drive the expression of higher levels of EGL-1 (61). For example, in the NSM lineage, *egl-1* transcription in the NSM sister cell is driven by HLH-2/3 and is inhibited in the NSM neuron by CES-1. Survival of hermaphrodite specific neuron (HSN) in hermaphrodites is driven by binding of TRA-1 to a distal enhancer (62). Once EGL-1 protein reaches a critical threshold, the anti-apoptotic interaction between BCL-2 ortholog CED-9 and Apaf1-like adaptor CED-4 at the outer mitochondrial membrane is disrupted by a physical interaction between CED-9 and EGL-1. Once free from CED-9, two isomers of free CED-4 (4 sets of CED-4S and CED-4L) form octamers that can bind to two molecules of the zymogen CED-3 (essentially forming two hetero-tetramers) where cleavage of a specific residue activates CED-3 (aCED-3), the main caspase responsible for physically remodeling/dismantling. After specification and execution phases, the dying cell is engulfed by one of its neighbors. aCED-3 also activates CED-8, a lipid scramblase that is important for exposing phosphatidylserine (PS) groups on the outside of the dying cell (63). The exposure and concentration of

PS trigger engulfment of the dead cell via CED-1, a MEG10 (multiple EGF-like domains-10) single-pass transmembrane receptor protein exposed on the membrane of the engulfing cell. This interaction leads to receptor/signal clustering, which activates two parallel engulfment pathways (reviewed in (59)).

Caspase-dependent PCD

Caspases are a family of cysteine proteases that are crucial for apoptosis. While CED-3 is the main executioner caspase in *C. elegans*, flies and mammals have multiple caspases that can be split into two groups: initiators and executioners. Initiator caspases have a long N-terminal pro-domain that mediates formation of protein-complexes for caspase activation. Executioner caspases are maintained in the cytosol as zymogens, but separation of its two subunits (p10 and p20), in addition to removal of the N-terminal pro-domain activates executioner caspases (64). The main executioner caspase CED-3 can then coordinate dismantling the infrastructure of the cell which includes elimination of survival signaling and generation of engulfment signals for neighboring cells (56). While mammals such as mice and humans have 10 and 11 caspases respectively, *C. elegans* has one major caspase, CED-3, involved in PCD in both the germline and soma and is essential for embryonic PCD (64). Like mammalian caspases, CED-3 primarily exists as a cytosolic zymogen and once in an octomer complex with CED-4, activated CED-3 (aCED-3) in the soma cleaves specific proteins that initiate initiate nuclear fragmentation such as the ribonuclease DCR-1 (dicer 1, ribonuclease III), or engulfment in a neighboring cell.. When DCR-1 is cleaved by aCED-3, the C-terminal fragment contains RNase

domains and is able to make nicks in DNA. These nicks demarcate DNA for further fragmentation and clean-up by a multi-nuclease complex including (amongst others) mitochondrial endonuclease CPS-6 and worm apoptosis-inducing factor homolog WAH-1. Once released from mitochondria, WAH-1 can also bind to scramblase SCRM-1, which flips lipids to expose phosphatidylserine (PS) groups on the exterior of the cell, a signal for engulfment. Additionally, aCED-3 can also cleave CED-8, a homolog of Xk-related transporter proteins whose c-terminal fragment promotes PS-externalization, the anti-apoptotic or pro-survival Bcl-2 homolog CED-9 (65, 66), in addition to inhibiting growth and proliferation signaling via the phosphoinositide 3-kinase (PI3K)/AKT signaling pathway by cleaving CNT-1, an ortholog of human ACAP1 and ACAP2 (67, 68). Interestingly, Snail-like protein CES-1 potentially regulates asymmetric sorting of aCED-3 into the dying cell – where sister cells fated to die have been shown to be physically smaller and with higher levels of aCED-3 and EGL-1 in the NSM and Ql.p neuroblast lineages (69). Conradt *et al.* have also shown that this may occur via coordination of transcription factors *hlh-2* and *hlh-3* and a gradient of aCED-3 is maintained by two engulfment pathways, ensuring that aCED-3 is asymmetrically dumped into the dying cell. A strong loss of function alleles of the gene *ced-3, n717*, causes the loss of asymmetric cell size (70, 71).

There are three additional caspase-like proteins, *csp-1*,*2* and *3* in *C. elegans*. aCED-3 can also autocatalytically activate itself, so another aspect of regulation is mediated by inhibiting aCED-3 from activating more zymogens of CED-3 (56, 72). CSP-2 and CSP-3 in *C. elegans* are similar to initiator caspases in mammals in that

they regulate the zymogen and subsequent activation of CED-3 in the germline and soma respectively (73, 74). However – CSP-1 has shown to be involved in CED-3-dependent and independent cell death in the touch neurons found in the anterior pharynx (72). Denning et al. (72) have found that CSP-1B and C are involved in embryonic PCD and expressed in the germline and maternally loaded during the pachytene stage of oogenesis. CSP-1B shown to have pro-apoptotic activity in PLM neurons in the anterior pharynx, while CSP-1A null mutations did not affect PCD. However – it is possible that CSP-1A may still be involved in apoptosis, but may be cell-specific, i.e. dependent upon lineage-specific gene expression. Interestingly, cell death involving CSP-1B is independent of aCED-3 and CED-4, suggesting that other pathways or factors in parallel to the canonical cell death pathway may be involved.

Left-right asymmetry of gut-gonad orientation

C. elegans is an established model organism for studying different methods of L/R cell-fate specification. It has been previously shown that the major structural chirality of the *C. elegans* adult is established during the division of the two anterior granddaughters of the zygote at the 4-6 cell stage of embryogenesis (Figure 4.8.) (75). Anteroposterior skewing of mitotic spindles occurs with a defined handedness driven by actomyosin-directed chiral cortical flows that generate torque as these cells divide (76) - similar to working models of the early vertebrate L/R symmetry breaking event involving Nodal (77). However, many L/R asymmetries that are established later in development require separate mechanisms to ensure cell-fate, such as the

asymmetric specification of AWC neurons or Notch signaling in early gut development (78, 79). The most prominent example of this L/R asymmetry in the worm is the orientation of internal organs - the gut and gonad. During embryogenesis, gut cells undergo stereotyped movements that require asymmetric Notch signaling (80): It has been hypothesized that orientation of the intestine determines the orientation of the gonad primordium cells Z1-Z4 and their descendants (81). In hermaphrodites, Z1 and Z4 become distal tip cells that lead the migrating gonad arms towards the anterior and posterior ends of the animal in order to form their characteristic J-shape. In males, Z4 becomes a distal tip cell and remains near the middle of the animal, while Z1 becomes the Linker Cell. The Linker cell leads the single gonad arm of the male and dies after connecting the vas deferens to the cloaca (82). It is thought that gut-gonad internal asymmetry depends on the skew or particular orientation of the descendants of Z1 and Z4 relative to the intestine, in addition to guidance cues such as Netrin signaling that help lead Z1 and Z4 (83). However – the GRN or gene that regulates this asymmetry is unresolved. While we or others have never seen a reversed N2 animal of either sex at standard conditions (20˚C) (75, 84, 85), we have reported that some N2 males (~5%) have reversed gut-gonad orientation independent of the left-right asymmetry established during early embryogenesis at increased temperature (25˚C). This was indicated by lack of reversal seen in other internal asymmetries of the nervous system such as *gcy-5* expression that marks the ASER neuron and coelomocyte position (another marker of internal asymmetry) (86). We have also found significant variation in the propensity for gut-gonad reversal frequency of males in 100 *C. elegans* isolates. Correlation analysis using ~40 isotypes suggest that gut-

gonad reversal is independent of migration errors of the gonad-arm, suggesting that Netrin signaling, or other guidance cues may not entirely be responsible for heterotaxy in males. To identify causal variants, we used association mapping and found phenotypic variation is associated with QTL on chromosomes II and III but further analysis of near-isogenic lines to identify causal variants is required (Chapter 4, originally published here (83, 86–88)).

Aims of this thesis

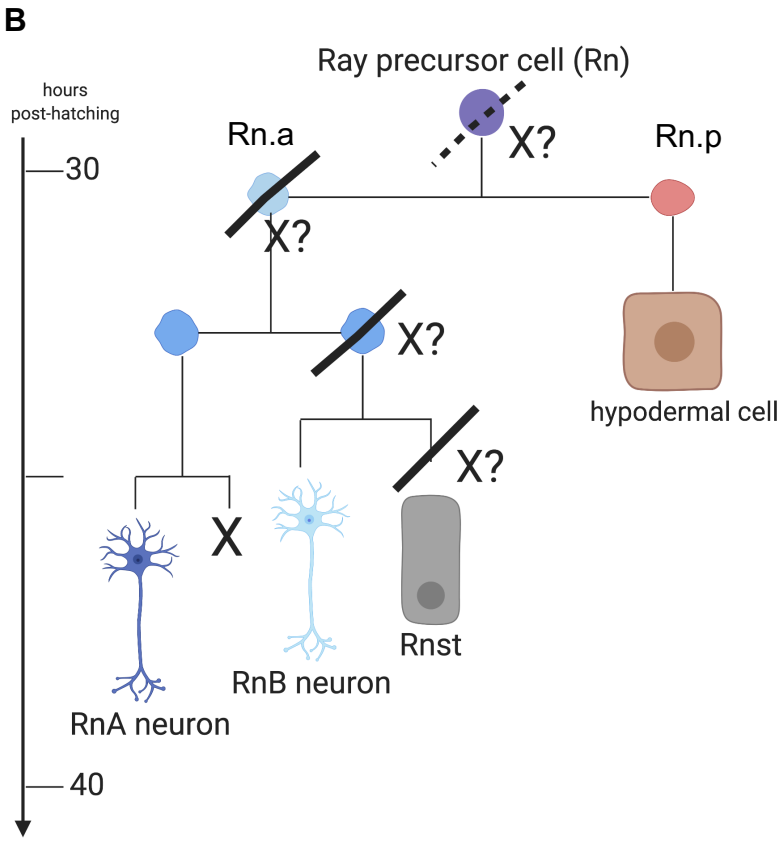
Considering the number of events that need to occur for successful development, it is remarkable that it happens correctly so frequently in any species. It is thought that coordination of specific gene regulatory networks is essential for proper development to occur. Wiring diagrams describe the inputs and relationships between genes such as transcription factors and their downstream components. However, they do not capture any naturally occurring variation, which is likely to occur in even the most canalized or robust systems (1, 14, 26, 27). Here we use the well-described lineage in a highly reproducible, robust model system of *C. elegans* to identify genetic components that regulate development of two processes: sensory ray development of the male tail and left-right orientation of the gut and gonad. Variation for both traits in closely related species suggest that the variation in N2 is not entirely due to lab domestication or hermaphroditism and there is some plasticity in associated GRN(s). We use two association mapping methods and the genetic diversity of n > 85 *C. elegans* isotypes (number varies per trait, see Chapters 3-5) to identify multiple loci for both phenotypes. In chapter 2, we show that stochastic ray loss shows a right-

hand bias and is due to programmed cell death which is completely suppressed by loss of the pro-apoptotic factor EGL-1. We also show that while majority of these EGL-1 dependent cell deaths in the male tail are dependent on the canonical PCD caspases- CSP-1 and CED-3, a small percentage of the EGL-1 dependent ray losses occur even in absence of the caspases. These residual defects in these mutant backgrounds suggest other genes are involved. While there is shared QTL for both traits on chromosomes II and III, these regions are large and contain many genes and does not necessarily indicate that genetic regulation of regulation of both processes is shared (Chapters 3 and 4). Analysis of near isogenic lines to study the independent effect of each loci *in vivo* excluded a small region on chromosome II and variant effect prediction software suggests many interesting candidates - however further work is required to narrow regions and identify causal variants. Additionally, we have been working on a parallel method for association mapping using ElasticNet regression, a multivariate method used in machine learning. We used ElasticNet regression to identify genes associated with the naturally varying requirement for the maternal transcription factor SKN-1 in endoderm formation in *C. elegans* isotypes. While it successfully identified previously associated loci that have been confirmed *in vivo* with near isogenic lines, it also identified novel loci on chromosomes V and X. Further analysis of the QTLs on chromosomes V and X with additional near isogenic lines is required to test if these regions are spurious associations due to population structure which is not accounted for by the ElasticNet pipeline. Analysis of gut development in near-isogenic lines and in strains carrying deficiencies encompassing these loci would

allow to narrow these regions to find a causal variant and will be informative as to the effectiveness of current methods.

Figure 1.1. Terminal sub-lineage of sensory ray development in male tail development.
(A) DIC overlay of the posterior portion of an N2 male expressing RAM-5::GFP, a marker for structural support cell (Rnst). 40X dorsal view, anterior to the right. Dendritic endings of neurons are located within the rays, while cell-bodies are found within the tail (89). Photo courtesy of Caroline Ackley. (B). A modified diagram of sensory ray terminal lineage based on the lineage originally annotated by Sulston (49, 90). Timescale in hours-post-hatching are marked on the left. Seam cells V5, V6 and T (L/R) (not pictured) divide at the late L3 stage to produce 18 independent ray precursor cells with same terminal sub-lineage: the posterior daughter-cell of Rn (red cell, Rn.p) produces a hypodermal cell while the anterior daughter cell Rn.a ultimately produces two motor-sensory neurons RnA and RnB (dark and light blue neurons) and a structural support cell (Rnst, grey cell), where the sister-cell of the RnA neuron is fated to die (marked by an X). We hypothesized that if apoptosis were to occur in other cells, such as the ray precursory cell (purple cell/Rn, black-line), this could phenotypically result in could result ray-loss, while loss of other cells such as Rn.a (light-blue cell) or neuroblasts that produce the structural support cell could result in errors in fusion with hyp7 or male SET cells during L4 stage (dark-blue cells, dashed lines). Lineage analysis diagram based on (54) and figure was created with BioRender.com.


A


B
Ray precursor cell (Rn)
hours
post-hatching
30
Rn.a
X?
Rn.p
X?
X?
hypodermal cell
X?
X
Rnst
RnB neuron
RnA neuron
40

Chapter Two

A novel role for CSP-1 in L-R asymmetry of distinct apoptotic pathways in *C. elegans* male tail development

Summary

While *C. elegans* somatic cell lineage is largely invariant, the male tail, which is important for copulation, shows a significant variation in the final number of the neuroepithelial structures called sensory rays. We present evidence for programmed cell death as the basis for this stochastic ray loss. Stochastic ray loss is regulated most strongly by the canonical PCD activator EGL-1. Approximately 39% of males in the laboratory reference strain of *C. elegans* (N2) have less than 18 (wild-type) number of sensory rays while 96% of ray defects are nearly completely suppressed in an *egl-1* (homologous to BH3-only proteins) null genetic background. Loss of other canonical PCD components including Apoptotic protease activating factor-1 (Apaf-1) ortholog CED-4 and the caspases CSP-1 and CED-3 also suppress missing rays defect albeit to a lesser extent. Additionally, we found asymmetries in ray loss defects; rays on the right-hand side are disproportionately affected in N2 and this phenomenon is independent of the major left-right asymmetry determining event that occurs at the 4-6 cell stage of embryogenesis. Interestingly, while *ced-3* mutants exhibit a right-hand ray loss bias, this bias is lost when all caspase activity is abrogated in null genetic background of the *csp-1; ced-3* double mutant. Ray loss may be a consequence of relaxed selection due to hermaphroditism or lab domestication. However, we have found that males from closely-related gonochoristic and hermaphroditic species had a varying propensity for ray defects and while the types of defects appear identical, i.e. losses or fusions, the specific rays affected varies between species as well, suggesting multiple genetic variants affecting the propensity for stochastic cell death in *Caenorhabditis* nematodes.

Introduction

Programmed cell death (PCD; apoptosis) is an evolutionarily conserved process in nearly all Metazoans that is essential for normal development and plays vital roles in tissue morphogenesis, termination of cell-lineages, and maintaining homeostasis (91). Regulation of PCD regulation is critical as excessive activation or misspecification can lead to widespread inappropriate cell death common to many types of cancers, certain autoimmune disorders and neurodegenerative diseases (17, 92–95). While transient or persistent sublethal activation of the PCD pathway can also lead to secondary apoptosis or chronic inflammation often seen in patients suffering from traumatic brain injury (96, 97), some cancers have evolved to escape cell death by upregulating pro-survival pathways, including–stress pathways, by taking advantage of the growth factors released by dying cells in tissues where other cells can divide to replace them which allows tumors to either regrow or evade chemotherapeutic drugs (98). Additionally, stochastic or dysregulated apoptosis and caspase activity has been thought to be the cause for decrease in overall cortical volume seen in brains of human patients with Schizophrenia, which is typically considered a neurodevelopmental and not necessarily neurodegenerative condition. fMRI studies have identified significant variation in regions of the brain affected by loss in cortical volume (18, 22, 99) and Karrer et al. (20) have modeled the most likely areas of the brain to be affected based on symptoms, but treatment with caspase-inhibitors alone has been mildly successful in only a subset of patients (100).

It is therefore critical for the viability of the organism to regulate PCD machinery - particularly caspase activity (see Chapter 1). However, the genetic and molecular mechanisms that regulate activity are extremely complex, and working with a model organism where cell-death mutants are viable and lack the genetic redundancy of mammalian systems (both flies and mice have multiple caspases) would be the most advantageous (57, 59, 101). Here we use *C. elegans* male tail as a model for stochastic or dysregulated cell death – specifically in postembryonic neurodevelopment of the male tail. *C. elegans* is a well-established model organism for genetic regulation of programmed cell death: the core apoptotic machinery is evolutionarily conserved, and the highly reproducible lineage provides a robust developmental system for precise quantification of variation within large isogenic populations (49, 50, 59). Electron micrograph (EM) data reveals that the entire nervous system of males has a total of 385 neurons, 155 muscles, and 39 non-muscle end organs, and the neural network that essential for copulation is comprised of 85 male-specific neurons with 64 sex-shared neurons. While the *C. elegans* may be small, it is still capable of complex, innate or genetically encoded mating behavior.

Typically adult males have 18 bilaterally symmetric neuroepithelial structures called sensory rays held within a cuticular fan and interstitial tissue (47, 48, 102). The neurons encased in rays relay sensory information from the moving hermaphrodite so that the male can orient itself and respond quickly by turning, posturing the tail, etc. (Figure 1.1) (51). The 9 pairs of bilaterally symmetric rays come from (presumably) identical hypodermal seam cells, V5 (L/R), V6(L/R) and T (L/R) and during the late

L4 stage. Each ray has an identical terminal sub-lineage where ray-blast cells (RnP) divide to create two motor or motor-sensory neurons, RnA and RnB, a structural support cell and a hypodermal cell. While post-embryonic lineage analysis demonstrates only one programmed cell death event per sensory ray, in agreement with Sulston's original observations during lineage analysis (49, 90), we found that ~39% of adult males have less than 18 sensory rays, where either one is missing, or two rays appear fused together (Figure 2.1). We hypothesized that potentially stochastic inappropriate cell death was occurring in the terminal ray-lineage, and if this were true then ray defects would be reduced in PCD-null genetic backgrounds. We found that ray defects were most significantly reduced in *egl-1*, followed by *ced-4* and *ced-3* –null mutants which are the conserved pro-apoptotic functional orthologs in *C. elegans* (64). Because we found that there were residual defects in both *ced-4* and *ced-3* null mutants, we reasoned that other caspases may be involved. CSP-1 is the only other caspase with known proteolytic activity with a mild effect on apoptosis during embryogenesis which is independent of CED-4 – the main CED-3 caspase activator. Here we found a novel role for CSP-1A and show that it has a small but significant reduction in ray defects when compared to N2 (Figure 2.2) but did not enhance the effect of *ced-3(n717) lof* mutants.

Additionally, we found that not all rays are likely to have a defect despite symmetrical lineages: rays 8 and 9, predominantly on the right-hand side of the tail, were disproportionately affected in the reference strain N2 and, most interestingly, also in a *ced-3(n717) lof* background. However, this asymmetry was not maintained

in animals where embryonic chirality was reversed or sinistral, as identified by orientation of the gut and gonad in *gpa-16(lof)* background (84) (Figures 2.3, 2.4). Neither sinistral nor dextral animals in a *gpa-16(lof)* background had a significant reduction in defects when compared to N2, suggesting that mis-regulation of PCD machinery is at the cellular level independent of lineages established during the 4-6 cell stage. Interestingly, the right-handed bias for defects is also no longer seen in *csp-1(tm917)II; ced-3(n717)IV* mutants or in males from closely related gonochoristic or hermaphroditic species, suggesting that there is standing genetic variation that may not be only due to lab domestication or any difference in selective pressure of males (Figure 2.4, 2.5).

Results

Stochastic ray loss requires pro-apoptotic factors

The male tail has 9 bilaterally symmetric pairs of neuronal sensilla (rays) arranged in a cellular cuticular fan. While the developmental program of *C. elegans* is executed with high-fidelity, a notable exception is the male tail which shows a high variation in the final number of rays with the terminal rays 8 and 9 being the most affected (49, 55). In agreement with Sulston (49, 55), we found that ~39% of adult males in the N2 reference strain (38.73%, n = 153/395, Figure 2.2) have less than 18 rays. Each ray consists of two neurons and a support cell, all of which arise from a common grandparent by a stereotyped lineage. It was suggested that the missing rays might result from the failure of the corresponding structural cells (Rnst), which are required for ray formation, to extend proper processes. Missing rays may be

attributable to abnormal fusion of rays due to mispositioning of ray axons as seen in Semaphorin mutants or to ray loss (103–108). We classified two major types of defects as either losses or fusions based on ray morphology (Figure 2.1), with ray fusions appearing fatter and ray losses characterized by gaps between rays with normal morphology. The male tail lineage shows that there is one stereotypical PCD event in the terminal cell-divisions of all 18 independent ray precursor cells (RnP) from the V5 (L/R), V6 (L/R) and T (L/R) lateral seam cells. However - we found (in agreement with Sulston) that ~39% of adult males in the N2 reference strain (38.73%, n = 153/395, Figure 2.2) have less than 18 rays. We classified two major types of defects as either losses or fusions based on ray morphology (Figure 2.1). We hypothesized that if both types of defects were due to stochastic cell death occurring at potentially different points in the lineage, then both types of sensory ray defects would be suppressed in PCD-null mutant backgrounds. We examined this question by assessing the requirement of core apoptotic regulators for the observed ray defects. In *C. elegans* PCD, relatively high levels of pro-apoptotic BH3-only domain protein EGL-1 (compared to cells fated to survive) initiates cell death by disrupting the physical interaction between CED-9 (a Bcl-2 homolog) and CED-4 (an Apaf-1 functional ortholog) that are localized in the outer mitochondrial membrane. Free CED-4 then activates executioner caspase CED-3, which can then initiate downstream signaling cascades for DNA degradation, inhibition of pro-survival signaling and promotion of engulfment by neighboring cells (48). We found that both types of defects (overall defect frequency) were decreased most significantly in *egl-1(n1062n1052)* (3.36% defects, n = 19/566, Fisher's Exact Test, 5% FDR, p =

$2.72x10^{-46}$), *ced-4(n1162)* (6.96% defects, n = 45/647, p = $1.21x10^{-35}$, 5% FDR), *ced-3(n717)* (15.59% defects, n = 97/622, p = $3.15x10^{-16}$, 5% FDR (44)), and *ced-1(e1735)I; ced-4(n1162)III; ced-3 (n717)IV* mutants (6.21%, n = 19/306, 5% FDR, p = $4.33x10^{-25}$) mutants when compared to N2 (Fig. 2.2.). These findings imply that the loss of rays inherent to development of the reference strain N2 is, at least in part, the result of stochastic apoptotic cell death in the ray lineages. However, males mutant for *ced-3(n717), ced-4*(*n1162)* and *ced-1(e1735)I; ced-4(n1162)III; ced-3 (n717)IV* genes still lack rays, suggesting that some rays are either lost by a non-cell death process or by a cell death program independently of *ced-3* or *ced-4*. One such candidate is CSP-1, an ortholog of human Caspase 2, as two of its isoforms CSP-1B and C have been shown to regulate PCD in the embryo but also remain independent of CED-4 and CED-9 (72). We found that a null allele of *csp-1 (csp-1 (tm917)II*; 750bp deletion in the coding region that removes the pro-domain of CSP-1A had a mild reduction in defects compared to N2 (26.55%, n = 30/113, 5% FDR, Fisher's exact test, p = 0.0244), but did not enhance the reduction of ray defects in *ced-3 (n717)* background (*csp-1(tm917)II; ced-3(n717)IV* males: 8.04%, n = 9/112, 5% FDR, p = 0.1954). We note that *tm917* removes the pro-domain of *csp-1A*, one of the three isoforms of *csp-1*. We will test caspases II and II, in addition to more alleles of CSP-1 such as *n4967* and *n5133* that eliminate the active site (52) to determine if CSP-1A has some caspase activity *in vivo* similar to activated CED-3 or if autoactivation of CED-3 is inhibited by an interaction of CSP-1A and activated CED-3, similar to CSP-2 and 3 in C. elegans.

Ray loss is independent of anatomical handedness

Given that V5, V6 and T left-right blast cells produce symmetric lineages wherein each terminal ray sub-lineage (n=18) has one programmed cell death event, we would expect that rays from either side of the tail would be equally likely to have an error if these errors were truly stochastic. To account for bilateral defects, we added the number of bilateral defects to both left and right number of defects to create a binomial variable and used one-sample binomial tests with re-sampling (n = 10,000 per genotype) to calculate 95% confidence intervals for the true expected proportion of right-handed defects for all observed genotypes. After correcting for multiple comparisons (5% FDR (85)), we found that there was a strong right-handed bias in N2 (n = 115/173, 95% CI: 0.51 < 0.6647 < 0.7345, p =1.926x10^{-4}) and c*ed-3(n717)IV* mutants (n=67/104, 95% CI: 0.5443 < 0.6443 < 0.7357, p = 0.02328) (Figure 2.3). We have also found that not all rays are affected equally - rays 8 and 9 account for 95.80% (n=137/143) of defects in N2 with rare events occurring rays 1-3 (Figure 2.1.). These contrast, for example, with *rnt-1* mutants, in which male rays are missing owing to improper specification of ray progenitors: in these mutants, R8 and R9 rays were never reported to be absent (84). While this was unexpected, especially considering that ray lineages from V5 (L/R), V6 (L/R) and T (L/R) are identical, we then hypothesized that potentially reversing the overall handedness of the animal would proportionally shift the right-hand bias seen in N2 or *ced-3(n717)IV* mutants. To reverse anatomical handedness, we used a temperature sensitive null allele of *gpa-16,* an RHO GTPase known to orient mitotic spindles during the 4-6 cell stage of embryogenesis. At standard growth conditions, ~20% of animals have completely

reversed handedness - indicated by gut-gonad orientation (86). While the overall propensity for defects did not decrease in *gpa-16(it143)I* mutants, we did find that the natural bias seen in N2 and *ced-3 (n717)IV* mutants was no longer maintained (Figure 2.3). Interestingly, we also found the handedness in *ced-3(n717)* mutants was no longer present in *ced-3 (n717)IV;csp-1 (tm917)II* mutants but not *gpa-16(it143)I; ced-3(n717)IV* mutants, suggesting differential role of CSP-1 in left and right ray lineages and potentially a role for GPA-16 (independent of its role in early symmetry-breaking event).

Propensity for ray defects varies in closely related gonochoristic nematodes

It is not clear what effect, if any, lab domestication has had on this phenotype, or if it is a consequence of hermaphroditism (109, 110). Kiontke et al. (111, 112) have identified significant phenotypic diversity in the male tails of globally diverse, closely related *Caenorhabditis* nematodes – specifically in the shape of the spicule, fan and positioning of rays 4 and 5. We hypothesized that potentially the residual ray defects seen in canonical cell death caspase CED-3 mutants could be due to lab conditions or potentially because they are a hermaphroditic species and if this was the case, males from gonochoristic species would have fewer defects than hermaphroditic species. Generally, we found this trend that males from hermaphroditic species have significantly more defects than those from gonochoristic species; both *C. elegans* N2 and LSJ1 (two domesticated lines) have greater propensity for defects than *C. remanei*, *C. brenneri*, *C. japonica* and *C. portoensis* (Figure 3.1, N2: 36% (n= 90/250) *vs*. *C. remanei*: 13.19% (n = 12/91), Fisher's Exact test, 5% FDR, $p = 8.57 \times 10^{-5}$; *C. brenneri*: 4.17% (n = 1/24), 5% FDR, p = 0.00210; *C. japonica*: 0% (n=0/91), 5%

FDR, p = $4.76x10^{-14}$; *C. portoensis*: 0.901% (n=1/111), 5% FDR, p = $7.74x10^{-15}$) and (LSJ1: 26.41% (n=42/159) vs *C. remanei*, , 5% FDR, p =0.03142; *C. brenneri,*5% FDR, p = 0.3124; *C. japonica*, 5% FDR, p =$3.48x10^{-9}$; *C. portoensis*, 5% FDR, p = $2.81x10^{-9}$). And while the types of defects, such as fusions or losses were consistent throughout all species observed, N2 has a unique (thus far) right-handed bias for defects predominantly in ray-pairs 8 and 9 (Figure 2.4), *C. briggsae* has ray defects in rays 3 and 4, and all rays predominantly on the left seem to be equally affected in *C. remanei* (Figure 2.4). While this does not help to identify genetic loci, it does suggest that the residual cell death errors in N2 cell death null mutants may not be entirely due to lab domestication or selective pressures or a consequence of hermaphroditism. Additionally, while the overall propensity and specific rays affected varied across the genus, the source of developmental errors may be occurring at the intersection between ray identity and concurrent morphogenesis events.

Discussion

The connections made between neurons are essential for cellular function and will ultimately form the behavior of an organism. Programmed cell death plays many roles throughout neurodevelopment (such as synaptic pruning, quantitative matching, etc.; (91)), which is also evident in the neuronal AB lineage in *C. elegans*, a fundamental model to understanding PCD. However, they do not capture any naturally occurring variation that has been identified in N2 with high-resolution imaging techniques, such as in vulva cell fate patterning (113) or in this case, as Sulston originally noticed, the final number of sensory rays in the adult male tail. We have found significant variation

in the male tail sensory ray lineage that predominantly affects rays 8 and 9 from the T-lineage. Interestingly, we found that *egl-1(n1048n3082)V* had the greatest reduction in defects when compared to N2, followed by Apaf1 ortholog *ced-4*, and conserved executioner caspase *ced-3.* The remaining frequency of defects in these mutant backgrounds suggest that CED-3-independent cell death is occurring. While we have not looked at all combinations of CSP-1/2/3 mutants, we did find a small reduction in defects in a *csp-1(tm917)II* mutant background, suggesting that CSP-1A has a pro-apoptotic role in somatic apoptosis (73), but further mechanistic analysis is required (Figure 2.6).

All L-R asymmetries of determinate handedness in *C. elegans* are thought to arise from a single symmetry-breaking event: the skewing of mitotic spindles in the four-cell embryo, leading to a defined ("dextral") embryonic chirality. The resulting bilaterally asymmetric cell-cell contacts lead to L-R asymmetry in cell lineages, cell identities, and overall anatomy. However, unlike all L-R cell pairs known to be affected by this early symmetry break, the ray progenitors arise from lineages that are entirely bilaterally symmetric. We therefore assessed whether the L-R asymmetry in deployment of alternative PCD pathways depends on the chirality of the embryo by analyzing animals with reversed handedness. A fraction of *gpa-16* mutants show mirror-image handedness compared to normal animals, owing to reversal in the direction of skewing of transverse mitotic spindles in the early embryo at standard conditions. Of the ~ 39% of N2 males with missing or fused sensory rays, 95.80% of these defects occur in rays 8 and 9 and both N2 and *ced-3(n717)IV* null mutants also

have a significant right-hand bias. This bias is lost when either ray defects are strongly suppressed when the core cell death machinery is compromised by mutations in a *ced-4(n1162)III, ced-4(n1162)III; ced-3 (n717); ced-1e1375); or egl-1n1048n3082)V* null backgrounds, or in *gpa-16(it143)I* mutants regardless of their internal asymmetry. Additionally, *gpa-16it143)I;ced-3(n717)IV* mutants do not have significant reduction in ray defects when compared to *ced-3(n717)IV* but the right-handed bias is lost. This suggests that potentially these defects are independent of orientation of the internal organs and GPA-16 may be playing a role at the cellular level.

Diversification of neuronal identity relies on coordinating gene expression in the right developmental time and place (114–116) and it has been shown in the NSM lineage that the transcription factors CES-1/2 bind to *egl-1* locus to regulate its expression - where cells fated to die will have relatively high *egl-1* expression compared to its sister-cell (69) and modifications of the cytoskeleton may have an effect on cell-fate. However - null mutants for both caspases known to be active in embryonic PCD and their activator still have ray defects, which are nearly completely suppressed only in an *egl-1 (n1058n3082)* background, suggesting caspase-independent cell death - i.e. additional genetic factors, are involved in this process. An interesting candidate is CSP-3, an inhibitor of CED-3 in the soma or *sma-9* – a transcription factor that interacts with the worm ortholog of human Bone Morphogen Protein 10 (BMP10) through the DBL-1 pathway (117). It has been reported to reduce ray 8-9 fusions in N2, but the decrease from 49% to 23% is similar to the effect size of CSP-1A, and the reduction of ray defects seem to be stronger in mutants for PCD, especially those for

egl-1(n1048n3082)V and *ced-4(n1162)III* mutants, which suggests that other genes may directly regulate EGL-1 expression in or there is an intermediary or parallel pathway involving *sma-9* and *egl-1.*

Recently it has been shown that the CED-3 protein is somehow sorted into the dying sister cell in the NSM lineage (71) It could be that active CSP-1A may be helpful to preferentially sort activated CED-3 (aCED-3) into the dying sibling cell in the terminal ray lineage, but the mechanism and if any other gene(s) are involved remain to be determined. It is not clear how loss of rays, specifically on one side, may be evolutionarily favorable as laser ablation studies have shown that rays 8 and 9 are important for a specific component of mating behavior – the turn around the posterior or tail-end of the hermaphrodite and may be an adaptation to laboratory conditions (53). It is curious that N2 males also have a preferential turning bias while mating (118) and it may not be correct to consider ray loss as a "defect" as it has been suggested that brain lateralization offers different fitness advantages in other species (119–124). Regardless of whatever the evolutionary pressure might be, this variability suggests that the gene-regulatory networks that pattern the male tail nervous system may be more plastic than others, and future work will focus on identifying these genetic components.

Methods

Strains used

Except where noted, *C. elegans* isolates were maintained as described by Brenner (125) and scored at room temperature. For a list of *C. elegans* isolates used in this study, see table below.

Strain	Genotype	Availability
N2	*C. elegans*	CGC
FX917	*csp-1 (tm917)II*	upon request
MT1522	*ced-3 (n717)IV*	CGC
MT2547	*ced-4 (n1162)III*	CGC
CU809	*ced-1(e1735)I; ced-4(n1162)III; ced-3 (n717)IV*	CGC
MT8735	*egl-1 (n1048n3082)V*	CGC
BW1809	*gpa-16 (it143)I*	CGC
JR3191	*gpa-16(it143)I; ced-3 (n717)IV*	upon request
JR3283	*gpa-16(it143)I;ced-4 (n1162)IV*	upon request
JR3196	*csp-1(tm917)II; ced-3n717)IV*	upon request
KC69	*wxIs29(pRF4+pKS+T24C2SphIgfpNLS-) I; him-5 (e1490) V*	upon request

LSJ1	*C. elegans* – N2 grown only in liquid culture	CGC
CB5161	*C. brenneri*	CGC
DF5081	*C. japonica*	CGC
EG4788	*C. portoensis*	CGC
HK104	*C. briggsae*	CGC
JU724	*C. remanei*	CGC

Generation of male stocks

Male stocks were generated and maintained as previously described (86, 87, 118). Briefly, 20–30 L4 hermaphrodites were picked into 7% ethanol solution in microcentrifuge tubes and rotated for 45-60 minutes (126). Worms were collected by centrifugation at 2000 rpm for 2 minutes and then transferred to fresh nematode growth medium (NGM) plates seeded with *E. coli* strain OP50 and incubated at 20°C. F1 male progeny were mated with sibling hermaphrodites to establish male stocks.

Scoring male tails and imaging

To quantify ray defects, 20-30 young adult males (1-2 days old) were picked into 3.5µl 5mm levamisole (127) on a 5% agarose pad. Paralyzed males were rolled until ventral-side up using an eyelash tool (an eyelash glued to a tooth-pick similar to a very fine paintbrush) and slides were sealed with additional 7 µl liquid (ddH2O if levamisole has completely evaporated). Male tails were imaged at 60x using a Nikon X-Cite 120 Series or 63X using a Zeiss Axioskope II. The ray(s) affected, gut-gonad

orientation, and type of defect was recorded. We classified ray defects into two categories, losses and fusions. We categorized a ray loss as a discernable gap where the ray would be, and neighboring rays do not appear affected by the missing ray (i.e. if ray 8 is missing, then rays 7 and 9 are still in place). We considered a fusion to be when there are no discernable gaps between rays - i.e. if rays 8 and 9 are fused together, then there would be no gap between rays 7 and 8 or 8 and 9 (See figure 2). We also found bilateral defects where the left or right side would have two different defects, i.e. a fusion and loss, and the affected rays on each side would not always be the same. For example, rays 8 and 9 may be fused on the left but ray 8 may be missing on the right. To account for this, we counted 1 defect for ray 8 loss on the right plus 1 defect for the fusion on the left between rays 8 and 9. Then we added 0.5 for bilateral-loss and 0.5 for bilateral-fusion. Essentially, a bilateral defect where each side of the tail accounted for 3 defects instead of 2. Additionally, fluorescent images were edited in ImageJ.

Statistics

All statistics were calculated using R version 3.5.3 (76). Fisher's exact test (2x2) was used to compare all tail frequencies unless otherwise stated. P-values were adjusted for multiple comparisons using false-discovery rate (128). 95% confidence intervals were calculated using bootstrapping (129, 130) (sampling with replacement) and the binom.test function (131). Plots were generated in R using ggplot2, ggsci, and cowplot (132–134).

Figure 2.1. Ray defects include losses and fusions
(A-C) Most commonly seen defects in N2 using differential interference contrast (DIC) microscopy (all posterior, ventral view) at 100x (A, B) and 63X (C). (A) Reference: 18 bilaterally symmetric rays (R – L indicates right and left-hand side). (B) Fusions: there is no discernable gap between 7 and 8 or 8 and 9, also ray 8 appears to be ~2x the size of typical rays 7-9 (compared to A and C). (C) Losses: Ray 8 is missing from the right-side, while rays 7 and 9 appear unaffected compared to reference (A) on both the right and left-hand side. Losses were recorded separately from fusions because of the discernable gap between rays 7 and 8 and 8 and 9.

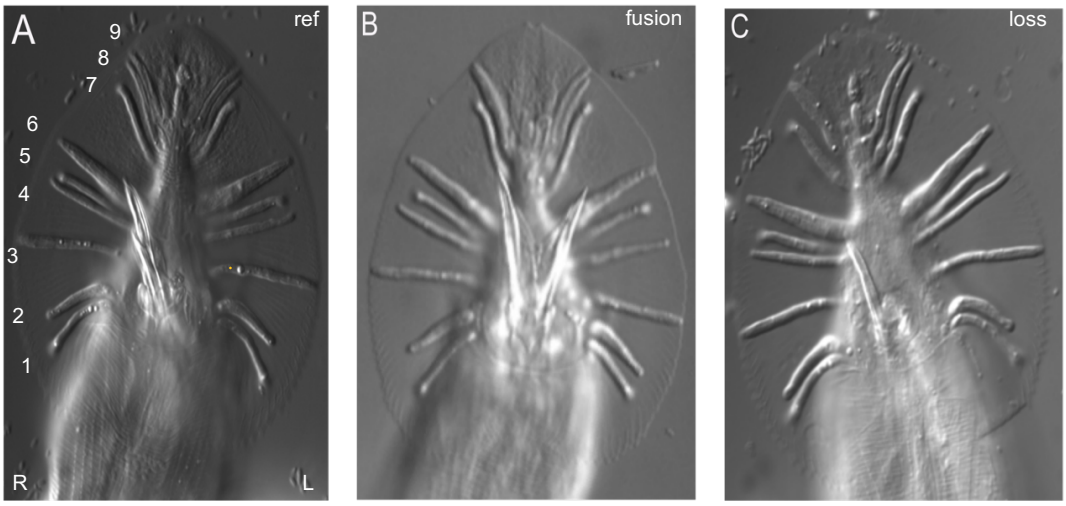

A
ref
9
8
7
6
5
4
3
2
1
R
L
B
fusion
C
loss

Figure 2.2. Propensity for stochastic cell death requires pro-apoptotic factors and is independent of anatomical handedness.
Defect frequency (including all defects – losses and fusions) is significantly reduced in PCD mutants. *it143* is a temperature sensitive allele of the gene *gpa-16* and at 20˚C, the majority of worms are phenotypically dextral while only ~ 20% of worms are sinistral. As a marker for internal asymmetry, the left-right orientation of the gut and gonad were recorded. All strains were compared against each other using Fisher's exact test. Multiple comparisons were corrected for using Benjamini and Hochberg method (= 5%) and significance level is depicted above each genotype. Significance codes: *p<0.05, **p <0.01, ***p<0.001, and no notation indicates p > 0.05.

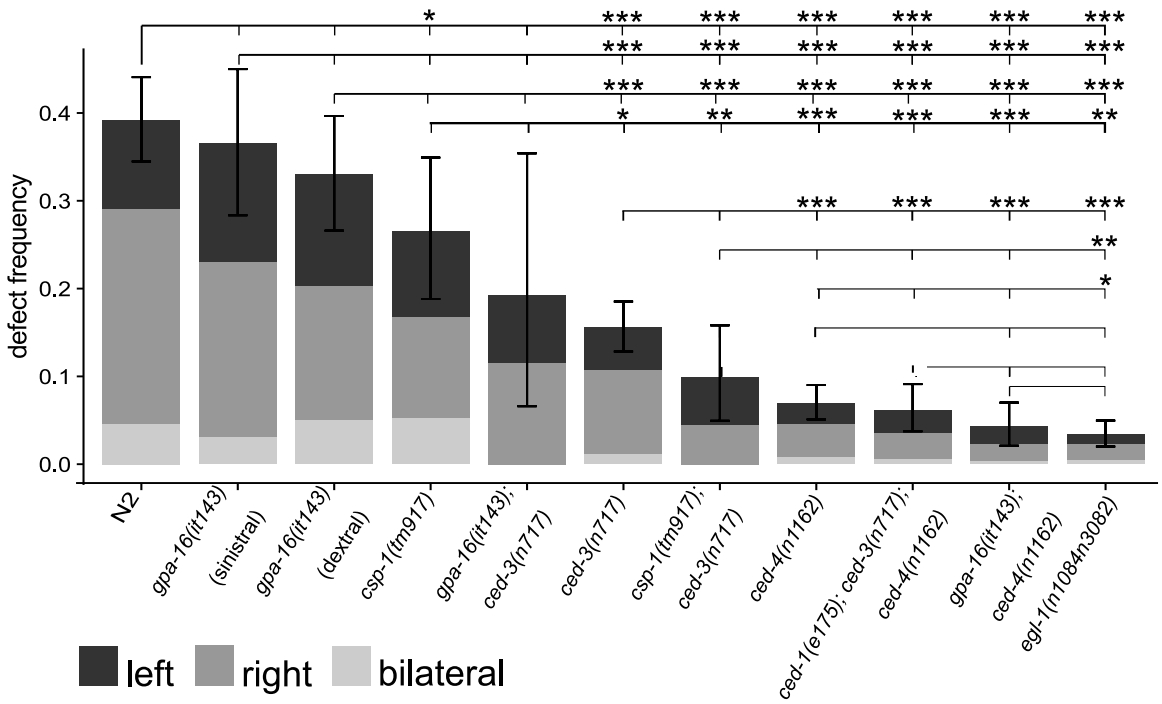

defect frequency
0.0
0.1
0.2
0.3
0.4
N2
gpa-16(it143)
(sinistral)
gpa-16(it143)
(dextral)
csp-1(tm917)
gpa-16(it143);
ced-3(n717)
ced-3(n717)
csp-1(tm917);
ced-3(n717)
ced-4(n1162)
ced-1(e175); ced-3(n717);
ced-4(n1162)
gpa-16(it143);
ced-4(n1162)
egl-1(n1084n3082)
left
right
bilateral

Figure 2.3. N2 and *ced-3(-)IV* males have a right-handed bias in stochastic ray-loss.

One-sample binomial tests for all genotypes with at least one defect at 20˚C. Given that V5 (L/R), V6 (L/R), T (L/R) blast cells produce identical lineages, we hypothesized that if strains had a propensity for defects then there should be no bias for the right or left-hand side (null hypothesis: p = 0.5). We added the number of bilateral losses and fusions to both left and right-hand defects to create two categories of defects. Both N2 and *ced-3 (n717)IV* mutants have a higher propensity for ray defects (mean = dots) on the right-hand of the tail, while the confidence intervals for other genotypes (including *gpa-16 (it143)* - sinistral animals) do not (95% CI exclude null hypothesis that rays from L or R blast cells would be equally likely to be affected by stochastic cell death). Significance codes: *p<0.05, **p <0.01 and no notation indicates p > 0.05. 20 < n <650 worms per trial, 1-3 replicates per genotype, all data is pooled. *n.b.* there are no chromosomes listed after gene names in figures for simplification, but these are mentioned throughout the text. Data also collected by Bilge Birsoy and Jo Downes.

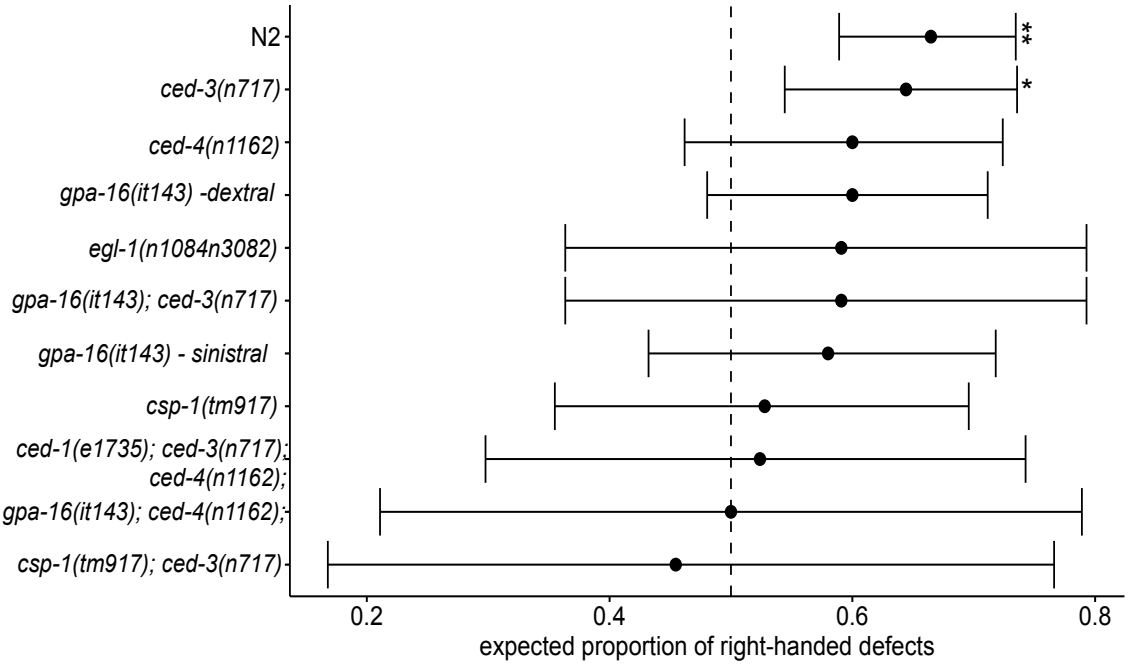

N2
ced-3(n717)
ced-4(n1162)
gpa-16(it143) -dextral
egl-1(n1084n3082)
gpa-16(it143); ced-3(n717)
gpa-16(it143) - sinistral
csp-1(tm917)
ced-1(e1735); ced-3(n717);
ced-4(n1162);
gpa-16(it143); ced-4(n1162);
csp-1(tm917); ced-3(n717)
**
*
0.2
0.4
0.6
0.8
expected proportion of right-handed defects

Figure 2.4. Overall propensity but not bias for defects varies in closely related hermaphroditic and gonochoristic species.

(A) Left, right and bilateral defects were counted in closely related hermaphroditic species *C. briggsae* and gonochoristic species (all else except for N2, LSJ1 and *C. briggsae*). The overall propensity for defects was compared against all strains and p-values were adjusted for multiple comparisons using Benjamini and Hochberg method with 5% FDR. (B) 1-sample binomial tests with resampling. The number of bilateral defects was added to both left and right-sided defects to create a binomial variable. $20 < n < 650$ worms per trial, 1-3 replicates per genotype, all data is pooled. *n.b.* there are no chromosomes listed after gene names in figures for simplification, but these are mentioned throughout the text. Data also collected by Bilge Birsoy and Jo Downes. P-values were adjusted using 5% FDR. (A and B)* $p<0.05$, 'ns' indicates $p > 0.05$.

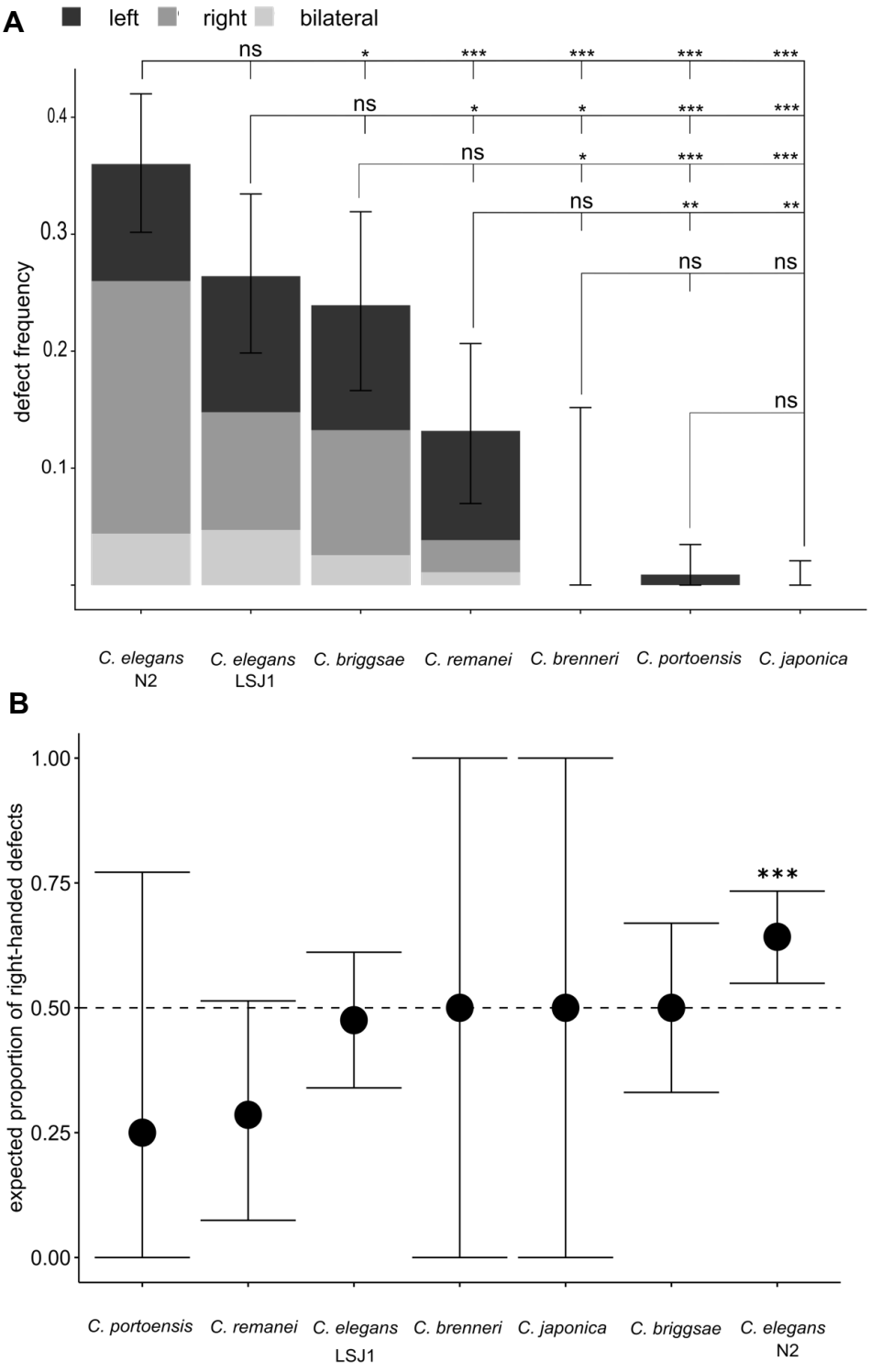

A
left
right
bilateral
defect frequency
0.4
0.3
0.2
0.1
ns
*
***
**
C. elegans N2
C. elegans LSJ1
C. briggsae
C. remanei
C. brenneri
C. portoensis
C. japonica
B
expected proportion of right-handed defects
1.00
0.75
0.50
0.25
0.00
***
C. portoensis
C. remanei
C. elegans LSJ1
C. brenneri
C. japonica
C. briggsae
C. elegans N2

Figure 2.5. Specific rays affected by stochastic cell death varies between closely related species.
For each defect, we recorded the specific ray(s) affected. Rays that appear fused together were counted as two separate defects.

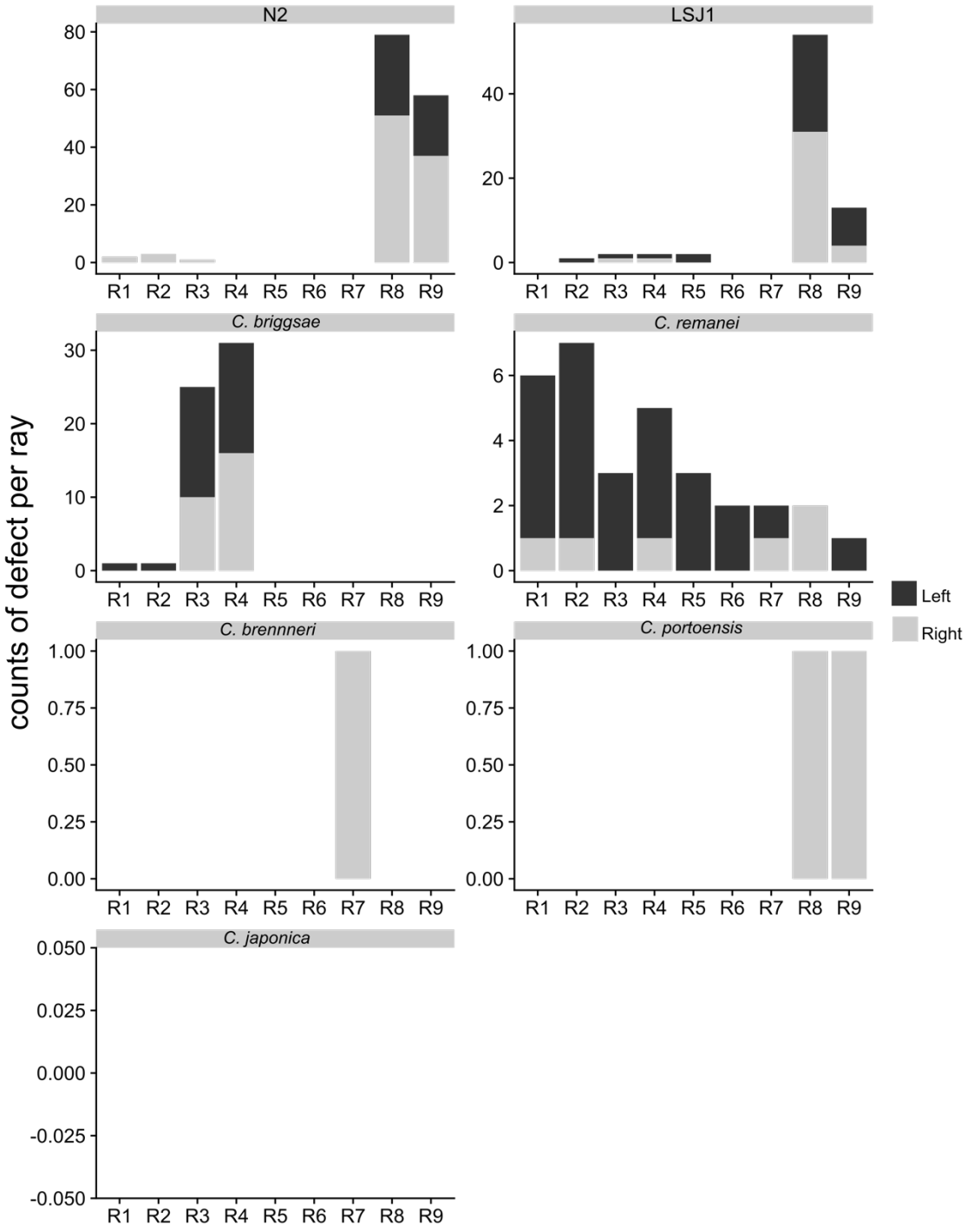
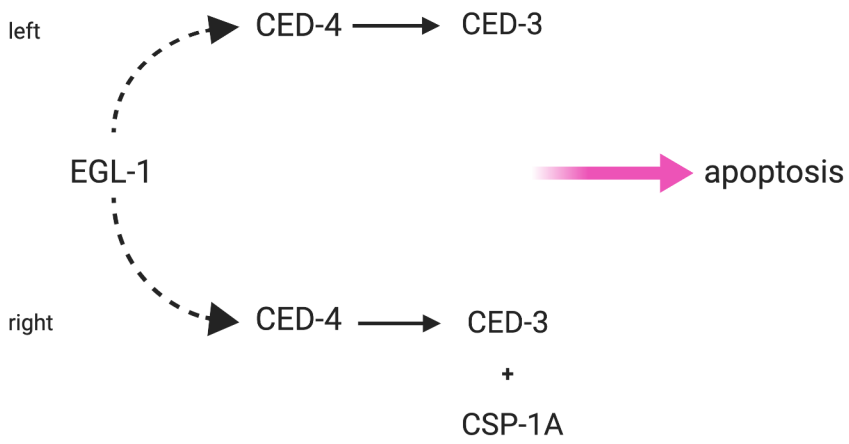

N2
LSJ1
C. briggsae
C. remanei
C. brennneri
C. portoensis
C. japonica
counts of defect per ray
R1 R2 R3 R4 R5 R6 R7 R8 R9
Left
Right

Figure 2.6. A novel role in regulation of apoptosis for caspase CSP-1 in male tail development of *C. elegans.*
Here we have evidence to suggest that missing or fused rays in the adult male tail are due to mis-regulation of apoptosis. We have found that males with strong loss-of-function (lof) alleles in PCD genes *egl-1, ced-3* and *ced-4* have significantly reduced defects compared to N2. Residual defects in these caspase and caspase-activator *lof* mutants suggest that other factors are involved in regulation of apoptosis – potentially downstream of *egl-1*. Here we show that the caspase 2 ortholog *csp-1* has a role in regulation of apoptosis, but how it is activated or specifically regulates apoptosis is to be determined and rays on the right-hand side are more likely to be affected in *csp-1* and *ced-3 lof* mutants. Figure based on Conradt et al. (59) and created with BioRender.com.


left
CED-4
CED-3
EGL-1
apoptosis
right
CED-4
CED-3
+
CSP-1A

Chapter Three

Propensity for stochastic programmed cell death naturally varies in male tail development of *C. elegans* isotypes

Summary
Previously we have shown that stochastic programmed cell death in the developing male tail in the *C. elegans* reference strain (N2) is highly dependent on the conserved programmed cell death regulator EGL-1*,* but not entirely on canonical caspases CED-3 or CSP-1, which suggests that other genetic factors are involved, and variation in positioning and final number of sensory rays also exists in closely related hermaphroditic and gonochoristic *Caenorhabditis* species, suggesting that stochastic ray-loss is not entirely due to lab domestication or relaxed selective pressure(s) that may exist in hermaphroditic species (Chapter 2). We have taken a quantitative approach to identify other genetic factors using the natural variation in the overall propensity for ray defects, which ranges from 0-~39% in 87 unique *C. elegans* isotypes. Additionally, *C. elegans* isotypes had the same types of defects in the same rays predominantly affected in N2 (Chapter 2) and knockdown of the pro-apoptotic gene *egl-1* via RNAi significantly reduced overall ray defects in most isotypes. We used two association methods to identify genetic variants: While GWAS identified 1 QTL on chromosome II using phenotypic variation of *C. elegans* isotypes, QTL-mapping with RILs made from isolates AB4 (~0.9%) and QX1211 (~20.2%) identified three independent QTLs that taken together account for ~55% of genetic variation associated with stochastic cell death. These results demonstrate there is potentially many QTL that could be affecting independent aspects such as the varied temperature response seen in *C. elegans* isotypes.

Introduction

Programmed cell death (PCD) is essential for neurodevelopment and plays a vital role in many processes including but not limited to tissue morphogenesis, termination of cell-lineages, and maintaining homeostasis (91). The same apoptotic machinery used during development also removes cells that have become deleterious to the organism and is the primary line of defense in cancers (135). Under stressful conditions, somatic cells may also undergo different types of death such as necrosis, autophagy or stochastic PCD (136). Stochastic PCD is typically activated by factors such as DNA damage, metabolic imbalances, errors in cell cycle, and accumulation of reactive oxygen species, as well as other forms of cellular stress. These cellular responses can be induced by changes in the environment, such as temperature and nutrient availability, but also perturbations by millions of years' worth of accumulated cryptic genetic variants. Robustness of development relies on the ability to withstand these different types of stress and deploy GRNs at the correct time and place, but cryptic genetic variants are also a source for necessary evolutionary change (7, 12, 13). The molecular mechanisms that regulate this balance are significant to a wide range of biomedical issues and fundamental questions in developmental biology.

Here we report significant phenotypic variation in the propensity for stochastic cell death during post-embryonic neurodevelopment of male tails in a globally and genetically diverse group of *C. elegans* isotypes. We propose this may be a model to identify GRNs associated with mis-regulation of somatic apoptosis. Building on Sulston's original observations (49, 90) of males with less than 18 rays during lineage

analysis, we have previously found that stochastic PCD during male-tail neurodevelopment is regulated by the evolutionarily conserved pro-apoptotic machinery EGL-1, CED-4, CED-3 and interestingly, an isomer of another caspase encoded by the gene *csp-1*. However residual defects in caspase and caspase-activator *lof* mutants suggest there are additional genetic factors involved in regulating this particular case of PCD. Further, we have previously found that the propensity and specific ray affected significantly varies between closely related hermaphroditic and gonochoristic species, suggesting that there may be cryptic genetic variants in addition to standing genetic variation in the gene regulatory network(s) that regulate PCD in sensory ray development (Chapter 2). The free-living nematode *C. elegans* is a leading model animal for studies of PCD (61, 63, 101, 137–139) and neurodegenerative diseases (140). This animal has a defined and reproducible number of cells: 959 somatic cells in the hermaphrodite and 1033 in males (49, 50) and the complete cell lineage, from single cell zygote, to adult in both sexes has been determined and found to be largely invariant from individual to individual. This allows for precise quantification of phenotypic variability in large isogenic populations and with an otherwise highly canonized development and life cycle.

Here we have quantified phenotypic variability in the development of the male copulatory structure and its associated neuroepithelial structures called sensory rays (as originally noted by Sulston (49)), which exhibit increased variability (reduced fidelity) in the total number of sensory rays. Ideally, *C. elegans* males will develop

eighteen bilaterally symmetric rays, where the dendritic processes of two neurons and a structural support cell together are wrapped in hypodermal tube (51–53). These rays are essential for copulation and we have found that in the laboratory reference strain of *C. elegans*, ~39% of males have missing rays grown at standard conditions (20˚C). We have also found standing genetic variation in the propensity for ray loss varies widely between 87 unique isotypes: while some virtually always have 18 rays, in others, up to approximately thirty-nine percent of males lack rays. Like the laboratory reference strain N2, the loss of rays in some isotypes appears to result from stochastic PCD as reduction of EGL-1, an evolutionarily conserved pro-apoptotic BH3-only domain protein, significantly decreased defects when isotypes were fed *egl-1* RNAi. We used quantitative trait locus (QTL) mapping with recombinant inbred lines (RILs (141)) made between two isolates: AB4 (which belongs to the CB4858 haplotype group), a strain with 0.9% defects and QX1211 – a strain with ~20% defects. The overall QTL model suggests that three independent loci found on chromosomes II, III and X accounts for ~55% of natural variation (ANOVA, *dof* = 4, $p = 3.2172 \times 10^{-11}$). We made several near isogenic lines between the two parental strains but were not able to identify any specific genetic candidates. However, analysis of some NILs and F1 males from reciprocal parental crosses did reveal epistatic relationships amongst the loci. While preliminary results from this system are limited so far, it could be a powerful model for quantifying the effect of genetic background in complex diseases.

Results

Ray loss varies across *C. elegans* isotypes

We measured ray defect frequency for 87 unique *C. elegans* isotypes (Figure 3.1) and calculated 95% binomial confidence intervals for each strain using Bayes method (142). Non overlapping CI suggest there is significant phenotypic variation, which ranges in *C.* 0.6-38.7% (Figure 3.1), which is similar to the range of phenotypic variation seen in closely related *Caenorhabditis* species (Chapter 2) (128). We also wanted to know if any of the other *C. elegans* isotypes had a bias for defects on the right-hand side of the tail similar to N2. Given there was at least one defect, we hypothesized that symmetric lineages from the L or R - V5, V6 or T lateral seam cells whose descendants include neuroepithelial sensory ray structures are independent and equally likely to have some kind of defect. To account for bilateral defects, we added the number of bilateral defects to the number of left and right defects to create a binomial variable. We calculated the probability that the proportion of right-handed defects in that specific ratio that would occur using bootstrapping and 95% confidence intervals. If lineages from left or right ray precursor cells are equally likely to be affected, we reasoned that these CI would include the expected proportion of right-handed defects of p = 0.5, and found no other isolates had significant bias for ray defects in the left or right ray lineages (Figure 3.1)

Reduction in EGL-1 suppresses ray defects

Previously we have shown that in the reference strain N2, males have missing or fused rays ~ 39% of the time and these defects are nearly completely suppressed in

egl-1(n1048n3082)(null) genetic background (Chapter 2). EGL-1 is the main pro-apoptotic regulator of programmed cell death in *C. elegans*, and we reasoned that since the types of ray defects in the different isotypes are phenotypically similar to N2, these errors may also depend on EGL-1 and therefore be due to stochastic cell death. To test this, we synchronized L1 populations of 6 different isotypes with a high propensity for defects and fed L1 larvae *egl-1* and L4440 (empty vector control) RNAi until they were approximately 1-2-day old adults (Figure 3.2). Ray defects were overall suppressed in N2 (10.97X reduction, 1.69%, n =1/59, Fisher's exact test, 5% FDR, p = 0.0203), ED3040 (2.20X reduction, 13.59%, n = 14/103, 5% FDR, p = 0.0203) and QX1211 (3.19X reduction, n=2/47, 5% FDR, p = 0.0203). While the other isotypes did not show a significant reduction in overall ray defects compared to control L4440 RNAi, they appear to follow the same trend as N2, ED3040 and QX1211 - as *egl-1* RNAi does modestly reduce defects (i.e. 95% binomial confidence intervals for JU1088, JU311 and CB4932 barely overlap, Figure 3.2.). These results suggest that ray defects observed in *C. elegans* wild isolates can, at least, be partly attributed to PCD.

GWAS identifies 1 quantitative trait locus (QTL) associated with stochastic ray-loss in *C. elegans* isotypes

To identify associated QTL, we used a genome-wide association method specifically designed for highly inbred model organisms (EMMA (143)) (Figure 3.3). We used a previously published SNP map (n = 4690 binary SNPs (144)) and linear mixed model analysis with an identity-by-state kinship matrix (n = 87 unique haplotypes/isotypes)) to identify any variants using the log10-transformed phenotype

(see materials and methods; this was to approximate assumptions of normality for linear-mixed models). After correcting for multiple comparisons using Bonferroni method (alpha = 0.05/ (4690 binary SNPs) = 1.066×10^{-5}), we identified seven SNPs that span a 1.85 mb region (~6.81 – 8.69 mb, 0.720 cM) on chromosome II (Figure 3.3). Effect plots (Figure 3.4) indicate that strains with major alleles (the most common at that particular locus) had higher defects than those with the minor allele at each SNP. While we found seven statistically significant SNPs, the possible region for a causal gene may be larger as it has been shown that the *C. elegans* genome has recently undergone extensive chromosomal sweeps where SNPs towards the center of the chromosome tend to be in high linkage-disequilibrium (LD) ($R^2 > 0.2$) (145). We used GCTA (146) to identify SNPs in significant LD ($p < 0.05$) with those identified by GWAS and found that these SNPs are in very high LD ($R^2 > 0.8$) with SNPs spanning a larger, ~3 mb region from 6.33 to 9.01 mb. (Figure 3.5 for a comparison of all regions identified by mapping techniques).

Standard interval mapping identifies 3 independent QTLs on chromosomes II, III and X

Previously, Baird *et al.* have used recombinant inbred lines (RILs) made between two isolates of *C. briggsae* that varied significantly in propensities for ray defects in rays 3 and 4 (1). They found that a significant percentage of ray fusions between 3 and 4 are regulated by *egl-5* (an ortholog of Human Homeobox A7 and B7 gene), suggesting genetic variation affecting stochastic cell death in male tail development is not restricted to caspases or known programmed cell death machinery (147). We

reasoned that the same strategy would work using *C. elegans* isotypes to potentially identify genetic variants and provide a larger picture of the gene regulatory networks and the heritability of these loci associated with stochastic cell death, especially because genome-wide association studies can miss loci with small effect sizes (148). Phenotypic variance from 74 RILs made from isolates AB4 (belonging to the CB4858 haplotype group (40), 0.9433%, n =1/106) and QX1211 (20.26%, n = 62/306) ranges from 0-35%, and comparison of ray defect frequencies among all RILs and parental strains suggest that this phenotype is only trending towards transgressive segregation (Figure 3.6). We used genotype-by-sequencing (GBS) to identify binary single-nucleotide-polymorphisms (SNPs) and R/qtl for standard interval mapping using Haley-Knott regression (see materials and methods). Standard interval mapping identified three independent QTLs on chromosomes II, III and X that accounts for ~55% of natural genetic variation for this phenotype (ANOVA, *dof* = 4, $p = 3.21 \times 10^{-11}$, Figure 3.5.). 95% Bayes confidence intervals span a ~6.5 mb region on chromosome II ranging from 3.85mb to 10.24 mb, an 11.5 mb region on chromosome III starting from the edge of the chromosome (~0.382 mb), and a region on X ranging from.18 mb to ~15.74 mb at a 5% genome-wide error level as determined by permutations (n = 10,000) (Figure 3.5, see overlapping regions). Effect plots of the SNP with greatest statistical significance from each region (determined by 95% Bayes confidence intervals) for chromosomes II and III predict that RILs with alleles from parent AB4 have a significantly lower propensity for defects which corroborates the region and pattern of alleles identified by GWAS (Figure 3.6).

Near isogenic lines exclude a small region on chromosome II

Because of recent chromosomal sweeps in the evolution of the *C. elegans* genome, it is not surprising that the QTLs identified are large relative to the length of the chromosome (145) (Figure 3.6). To narrow these regions and identify a causal variant(s), we constructed near-isogenic lines with small parts of QTLs on chromosomes II and X. While many lines were not viable, perhaps due to genomic incompatibilities between the two strains (115), we were able to make a NIL on chromosome II containing a small ~1.5 mb region (~1-2.71mb) from parent AB4 in a QX1211 background (Figure 3.8) and several NILs on chromosome X. Interestingly, we found phenotypic variants in spacing of rays 4 and 5 where these two rays appear like they are projected out from the same place or on top of each other in NILs on both chromosomes (Figure 3.8). Since we had not observed this type of defect in N2 or other *C. elegans* isotypes (Chapter 2), it is not clear if these should be considered part of the ray loss phenotype associated with stochastic PCD, and further phenotyping with *egl-1* RNAi may clarify this. If all defects are considered together, the small region of chromosome II did not suppress ray defects significantly compared to QX1211 (NIL II: 26.47%, n = 18/68, Fisher's exact test, 5% FDR, p = 0.3592, Figure 3.8.) but still had significantly more defects than AB4 (5% FDR, $p = 1.76 \times 10^{-6}$), which suggests the remaining region from ~2.77-10.38 mb may contain a causal variant(s) on chromosome II. We were also able to make two NILs that have QX1211 alleles from ~10.28mb to n < 14.25 mb (NIL XA-Lines 1 and 2, Figure 3.8.) or 11.53 < n < 14.25 mb (NIL XB, Figure 3.8.) in an assumed otherwise AB4 background. While NIL XB (12.77%, n = 6/47, Fisher's exact test, 5% FDR, p = 0.00902) and NIL XA-Line 1

(17.65%, n = 12/68, 5% FDR, p = 0.00414) had significantly more defects than AB4 but not QX1211, NIL XA-Line 2 showed the opposite behavior (NIL XA-Line 2: 4.92%, n = 3/61, p = 0.00870 when compared to QX1211). However - the region of interest on X determined by Sanger sequencing at specific SNPs suggests this region is the same as NIL-XA Line 1. Whole genome sequencing is required to confirm results from Sanger sequencing, identify more precise recombination breakpoints, and potentially other parts of the genome that despite 6 generations of backcrossing might still be from the opposite parental background QX1211. F1 males from reciprocal crosses showed the effects of chromosome X in isolation: males with X from AB4 and heterozygous on all other chromosomes (QX1211 males x AB4 hermaphrodites) have significantly more defects (n = 21/95) than F1 males from the reciprocal cross (6.73%, n = 7/104, Fisher's exact test, 5% False Discovery rate, p = 0.00727, Figure 3.8.). In addition, F1 males from QX111 males with AB4 hermaphrodites have a significant bias for defects to occur on the right-hand side of the tail (1-sample binomial test, 5% FDR, p = 0.02080) (Figure 3.8). While these results from these NILs are not conclusive, they suggest there are significant epistatic interactions.

Discussion

Precise regulation of apoptosis is vital for the overall health of an organism, as the same apoptotic machinery (such as caspases) have roles in other developmental processes such as axonal pruning and cell remodeling (57, 91, 135). While cellular processes like DNA replication, aminoacyl-tRNA charging, and repair of genome damage are considered high fidelity and have known error-correcting mechanisms,

not many genes, GRNs or mechanisms are known that may regulate or ensure precise spatiotemporal gene expression for successful development, particularly in times of cellular or environmental stress (3–5). One of the few, most well-known genetic mechanisms involves the chaperone protein Hsp90, where blocking its activity in wild or more domesticated lines of fruit flies produced variable, pleiotropic phenotypes not seen in either parent strain, suggesting it is capable of buffering noise that may occur from accumulation of random cryptic genetic variants for a wide variety of developmental processes (6–8). While homeostasis is essential for robust development, cryptic genetic variants are also material for evolutionary change and it is not clear how this balance is maintained (6, 7, 149). Here we exploit the predominantly determinate, invariant somatic development of *C. elegans* to identify genes associated with the propensity for errors – specifically in sensory ray development and report two major findings: 1) there is significant phenotypic variation within genetically diverse *C. elegans* isotypes in the final number of sensory rays in male tails; and 2) These ray defects may be regulated by evolutionarily conserved pro-apoptotic cell death protein EGL-1 – potentially occurring in different cells within the same lineage. 3) GWAS and QTL mapping with recombinant inbred lines identified several loci associated with stochastic ray loss, including cryptic X-linked genetic variants.

Variation in propensity for ray defects may be dependent on EGL-1

While the pattern of somatic PCD is largely invariant, we found that PCD errors occur in the male tail mating apparatus (Fig. 3.1.). A “normal” tail includes 18 sensory rays,

each with two neurons and a support cell. Rays are often missing in males of the laboratory N2 strain (49) and we found that mutants that do not undergo PCD (*ced-3*, *ced-4*, *egl-1*) suppress these losses (Chapter 2). Thus, ray loss appears to result from sporadic PCD "errors," typically not seen in other lineages (48, 49, 81). We found that the propensity for missing rays varies dramatically across 97 genotypically distinct wild isolates: while some isotypes, like N2, show high rates (nearly 40%) of loss, others virtually never (<1%) do (Figure 3.1); however, no other isotypes appear to have an asymmetric (left or right-handed) bias for defects. As in N2, losses in some strains with high propensity for ray defects are suppressed by *egl*-1 RNAi (Figure 3.2).

GWAS identified one QTL on chromosome II

Strains showing high and low fidelity in PCD provides genetic access to sporadic errors in PCD. GWAS and QTL analysis of RILs identified a shared QTL on chromosome II while QTL mapping indicated two additional QTLs on III and X (Figure 3.4, 3.5, 3.6). For association mapping, we used the standing genetic and phenotypic variation in *C. elegans* isotypes and a Bonferroni cut-off for significance (5% genome-wide error, alpha = 0.05/4690 = 1.066×10^{-5}). While it is the standard in quantitative genetics, it is quite conservative and there appeared to be SNPs very close to this significance threshold on chromosome III.

QTL mapping with RILs reveal additional loci on chromosomes III and X

We found significant genotypic and phenotypic variation in recombinant inbred lines. Effect plots of the most statistically significant SNPs identified by QTL mapping on

chromosomes II and III suggest that RILs with alleles from AB4 are more likely to have fewer defects than QX1211. However, the opposite relationship is predicted for SNPs on the X QTL. Analysis of from F1 males from reciprocal crosses suggest there are complex, epistatic interactions between X-linked and autosomal QTL, as F1 males from QX1211 males crossed with AB4 hermaphrodites revealed two effects: males had a significantly more defects than males from the reciprocal cross disproportionally on the right-hand side of the tail while males from the reciprocal cross had fewer defects and no bias. Given that the V5, V6 and T -L/R neuroblasts produce bilaterally symmetric rays that are identical in their distinct morphology, gene expression and function (55, 150), it was unexpected to see a bias for defects on the right-hand side of the tail in N2, let alone F1 males from parental isotypes with no right-handed bias. Because so many NILs were not viable after n < 6 generations of backcrossing, we used SnpEff to annotate variants found in genotype-by-sequencing data, filtered by location and predicted impact of the variant (Table 3.1.).

Robustness of development relies on precise deployment of GRNs. However, even in biological systems where their development is highly determinate and reproducible such as in *C. elegans*, we have found significant variation in fidelity of sensory ray morphogenesis in male tails. We have found a highly heritable GRN comprised of at least three QTL on three different chromosomes that regulate stochastic cell death male tail. We thus propose that *C. elegans* is an attractive model for testing the effect of SNPs not only on their associated phenotype but also on development as a whole

in different genetic backgrounds that regulate developmental PCD and asymmetry of cell death phenotype.

Methods

Key resources

Reagent Type (species) or resource	*Designation*	*Source or reference*	*Additional information*
C. elegans isotypes	See NAS for data	CGC	(1)
Caenorhabditis species	See NAS for data	CGC	
QX1211 x AB4 RILs		This study	Upon request
AB4 x QX1211 NILs	JR4018	This study	Upon request
QX1211* x AB4 *NIL	JR4019	This study	Upon request
Software	R v 3.5.1	The R Foundation	RRID:(22)
Software	PLINK	http://pngu.mgh.harvard.edu/purcell/plink/	RRID:SCR_001757

General maintenance of *Caenorhabditis* strains

All strains were maintained as described by Brenner (125) except where noted. All strains were maintained on NGM plates seeded with OP50 *E. coli* bacteria and kept at 20˚C.

Male Stocks

Briefly, male stocks were generated by aerating ~ 40 L4-stage hermaphrodites in 7% ethanol (400µl) for 45 minutes and maintained with standard conditions (126). F1 males were then mated to sibling hermaphrodites and their progeny was used for data collection.

Scoring male tail defects

1-2-day-old adults from male stocks (n = 20-30) are paralyzed in 3.5µl 5mM Levamisole (127). 5% agarose pads and worms are then oriented ventral-side up using an eyelash tool (single eyelash attached to a toothpick). 6 µl of M9 buffer (125) are added to a coverslip before covering the agar pad. Male tails were then scored using Differential Interphase Contrast (DIC) microscopy at 63x or 100x on a Zeiss Axioskope II or Nikon Ti Eclipse. The ray (s) affected, type of defect, (including left or right position of affected rays) and L/R gut-gonad orientation were then scored. We counted two separate classes of defects and organized them by position: ray-losses occur when there is a discernible gap between rays, or in the case of ray 9 missing then rays 7 and 8 appear to be the same size. This discerns a loss of ray 9 from what we classified as a ray-fusion: fusions occur when there is no discernable gap between

rays, i.e. ray 8 appears to be twice its normal size and there is no discernable gap between rays 7 and 8 or 8 and 9 (See Chapter 2).

RNAi

Males were allowed to mate with hermaphrodites for 24 hours on NGM plates. Plugged (43) adult hermaphrodites were picked into a 50:50 mixture of 1N NaOH: Bleach (20-30µl for 3-5 hermaphrodites). Embryos were then allowed to hatch overnight on unseeded NGM plates at 15˚C to generate synchronized L1 populations (151). Synchronized L1 were fed either *egl-1* or L4440 (empty control vector) RNAi at 20˚C and scored as 1-2-day old adults for ray defects. [egl-1 – ORF ID: F23B12.9 from Vidal Library]. RNAi bacterial strains from 80% glycerol stocks were streaked onto LB plates with 50mg/mL Ampicillin and 5 mg/mL Tetracycline and grown at 37˚C for 16-20 hours. A primary culture from a single colony was grown for 4-6 hours in 50mg/mL Ampicillin and 5 mg/ml Tetracycline. This culture is then diluted 1:10 and aerated for 4-6 hours at 37˚C. 1 M IPTG (1µl 1M IPTG for 100µl of primary overnight culture) is used to induce expression and 60 µl of induced secondary RNAi culture are seeded onto 35 mM plates containing 25 µg/ml carbenicillin and 1M IPTG. Plates are allowed to dry at room temperature (~22˚C) and used within 2-5 days (152).

GWAS

Genome-wide association study was performed in R with EMMA package (153). Ray defect frequencies from 87 unique isotypes were log-transformed for normalization. P-values were calculated using EMMA. REML.t () function and an identity-by-state

(IBS) matrix to account for population structure and assuming all SNPs have an additive relationship. We used a previously published SNP map with 4690 binary markers (38) and calculated 5% genome-wide significance threshold using Bon-Ferroni method where = 0.05/4690 = 1.066×10^{-5}. Generally, heritability (h^2) was calculated using the equation:

$$\frac{\sigma_g^2}{(\sigma_g^2 + \sigma_e^2)}$$

where h^2 equals the sum of genetic and were calculated by the emma.reml.t function per SNP. Narrow-sense heritability then equals the sum of genetic variance calculated by EMMA per statistically significant SNP – including those in significant linkage-disequilibrium – divided by the total genotypic and error-term variance (from all SNPs used for association mapping, n = 4690). Broad-sense heritability was calculated as the sum of genetic variance of all SNPs included in the model (n = 4690) divided by all genotypic and residual variance terms calculated by the EMMA package.

RILs and Genotype-by-sequencing

Recombinant Inbred Lines (RILs) were made by crossing QX1211 males with sperm-depleted hermaphrodites from isotype AB4 (141). 150 F2 progeny were singled onto separate plates and allowed to self for 10 generations by cloning 1 hermaphrodite per generation. A total of 84 lines were generated and immediately frozen at -80˚C. Genomic DNA was collected from 4 large populations of each RIL and both parents (4 x 100mm plates grown at 20˚C on OP50 and freshly starved) with a Blood and Tissue QIAGEN kit and modified protocol (154). GBS libraries and analysis were

prepped and analyzed as previously described (155). Briefly, ApeKI-digest fragments were selected by size (n < 100bp per fragment) and then used for library preparation. 84 RILs and both parental strains (only 74 were phenotyped) were sequenced using an Illumina HiSeq 2000 (2 lanes). Fastq files were aligned to reference genome WormBase 252 using BWA v. 0.7.8-r455 (156) and SNPs were filtered with vcftools (157) samples with more than 90% missing data and SNPs with a minor allele frequency (mAF) of < 1% were excluded from analysis, leaving 24,893 potential variants for analysis in R/qtl (142). SNP calling was performed using GBSversion3 pipeline in Trait Analysis by aSSociation, Evolution and Linkage (TASSEL) (158).

QTL mapping with R/qtl

Variants identified by Tassel 5 GBS pipeline were exported as an ABH genotype file (.csv). This was used to generate a cross object and genetic map in the R/qtl package containing1599 binary markers. The raw phenotype frequencies were transformed using square-root and normality was tested using Shapiro-Wilkes test. Standard interval mapping for single QTL analysis and permutations (n = 10,000) to determine genome-wide significance thresholds were calculated using the scaneone function. 95% confidence intervals for each region were calculated using the bayes.int function. Pairwise interactions between all possible QTL terms were calculated using scantwo function using standard interval mapping and imputation methods, and genome-wide significance threshold was calculated using scantwo function with stratified permutations (n = 1000). To estimate broad-sense heritability, we first used forward model selection using stepwiseqtl function and Haley-Knott regression. Model fit was

calculated using fitqtl function using drop-one analysis. 95% confidence intervals from the multi-QTL model were calculated using bayes.int function and analysis provided in (142).

Variant annotation using SnpEff

This fastq file was split into multiple fastq files with one sample using fastx toolkit and barcodes from sequencing (183). Once separated, we followed the first 7 steps of GATK best practices (112, 184). Each fastq file was aligned individually to the reference genome WS264 using bwa v.0.7.17-r1188 with the samse algorithm for mapping single-ended reads, and samtools was used to merge and sort all individual BAM files (185–188). GATK tools on the Galaxy platform was then used to mark PCR duplicates and then realign sequences around indels using basic default options. Variants were called using mpileup command from samtools and the call command from bcftools. Base-score recalibration was then performed and variants with Phred score less than the average 35 were removed. VEP was then used to annotate the resulting VCF using SnpEff v4.3 database: WBcel235.75.

Table 3.1. Candidate genes from VEP analysis of GBS data.
All candidate genes are in one of the associated QTL (either II, III or X) and were selected because they were predicted to have at least one change or variant that is predicted by SnpEff have a moderate effect. They also must fit the expected pattern if the QTL were independent, (see Figure 3.7 and spreadsheets on NAS drive for all genes). Additionally, based on phenotype data from WormMine (cite), known phenotypes for alleles of these genes include accumulated somatic cell corpses. Variants from our GBS data were also confirmed on CenDR (28)

Gene name	***Brief description from WormBase***	***Chr***	***Bp (start)***	***Bp (end)***
C27H5.4	Is an ortholog of human MFSD13A (major facilitator superfamily domain containing 13A). Is predicted to encode a protein with the following domain: Transmembrane protein 180.	II	7162444	7165849
cct-5	Is an ortholog of human CCT5 (chaperonin containing TCP1 subunit 5). Is predicted to have unfolded protein binding activity. Localizes to the chaperonin-containing T-complex.	III	4499187	4508121
mdt-15	Is an ortholog of human MED15 (mediator complex subunit 15). Exhibits nuclear hormone receptor binding activity and transcription coactivator activity. Is involved in… several processes, including determination of adult lifespan\; embryo development\; and locomotion.	III	5828557	5833745

	Is expressed in the head. Is used to study chromosome 22q11.2 deletion syndrome, distal.			
ubq-1	Is an ortholog of human UBC (ubiquitin C). Is predicted to have protein tag and ubiquitin protein ligase binding activity. Is expressed in the body wall musculature. Human ortholog (s) of this gene are implicated in isolated cleft palate.	III	5940244	5943835
rpn-2	Is an ortholog of human PSMD1 (proteasome 26S subunit, non-ATPase 1). Exhibits protein kinase C binding activity. Is predicted to contribute to endopeptidase activity. Localizes to several cellular components, including the cell cortex\; mitotic spindle\; and nuclear membrane.	III	6198902	6202626
C23G10.8	Is expressed in the tail. Is predicted to encode a protein with the following domains: NADAR and YbiA-like superfamily.	III	6202614	6206360
thoc-2	Is an ortholog of human THOC2 (THO complex 2). Is predicted to have mRNA binding activity.	III	6368042	6373904
copd-1	Is an ortholog of human ARCN1 (archain 1). Human ARCN1 exhibits RNA binding activity. Is predicted to encode a protein with the following domains: Longin-like domain superfamily\; AP complex, mu\/sigma subunit\; Coatomer delta subunit\; Mu homology domain\; and AP-2 complex subunit mu, C-terminal superfamily.	III	6622571	6624979

spk-1	Is an ortholog of human SRPK1 (SRSF protein kinase 1)\; SRPK2 (SRSF protein kinase 2)\; and SRPK3 (SRSF protein kinase 3). Exhibits protein serine\/threonine kinase activity. Is involved in embryo development and protein phosphorylation. Is expressed in several tissues including the gonad and the tail.	III	9456563	9463007
vha-1	Is an ortholog of human ATP6V0C (ATPase H+ transporting V0 subunit c). Is predicted to have proton-transporting ATPase activity, rotational mechanism. Is involved in several processes, including embryo development\; lysosomal lumen acidification\; and ovulation. Localizes to the membrane raft. Is expressed in the excretory cell and the rectum.	III	9781243	9782057
F26C11.3	Is expressed in the coelomocyte and the intestine. Is predicted to encode a protein with the following domains: Sushi\/SCR\/CCP domain and Sushi\/SCR\/CCP superfamily.	III	9903059	9915099
mua-3	Is an ortholog of human FBN1 (fibrillin 1)\; FBN2 (fibrillin 2)\; and FBN3 (fibrillin 3). Is predicted to have collagen binding activity and intermediate filament binding activity. Is involved in cell-matrix adhesion. Localizes to the hemidesmosome and intermediate filament. Is expressed in the epithelial system\; the muscle cell\; the rectum\; and the sensillum. Is used to	III	10160697	10183529

	study Marfan syndrome. Human ortholog (s) of this gene are implicated in Weill-Marchesani syndrome\; distal arthrogryposis\; and eye disease (multiple).			
sly-1	Is an ortholog of human SCFD1 (sec1 family domain containing 1). Is predicted to encode a protein with the following domains: Sec1-like protein\; Sec1-like, domain 2\; and Sec1-like superfamily.	III	10507058	10518726
tat-3	Is an ortholog of human ATP10A (ATPase phospholipid transporting 10A (putative))\; ATP10B (ATPase phospholipid transporting 10B (putative))\; and ATP10D (ATPase phospholipid transporting 10D (putative)). Is predicted to have phospholipid-translocating ATPase activity. Is involved in sterol metabolic process. Is expressed in several tissues, including the alimentary system\; the epithelial system\; the head\; the reproductive system\; and the tail.	III	10722147	10730705

Figure 3.1. Propensity for stochastic cell death varies in *C. elegans* isotypes. (A) left, right bilateral defects were measured and show 95% Bayes binomial confidence intervals for the overall proportion of defects. $98 < n < 400$ per isotype, 1 sample per isotype. (B) 95% binomial confidence intervals from 1-Sample binomial tests (with resampling) for expected proportion of right-handed defects. We added the number of bilateral defects to both the number of left and right defects to create a binomial variable. If rays from either left or right V5, V6 and T lineages are equally likely to have some type of defect, then these confidence intervals would include the expected proportion $\mu = 0.5$.

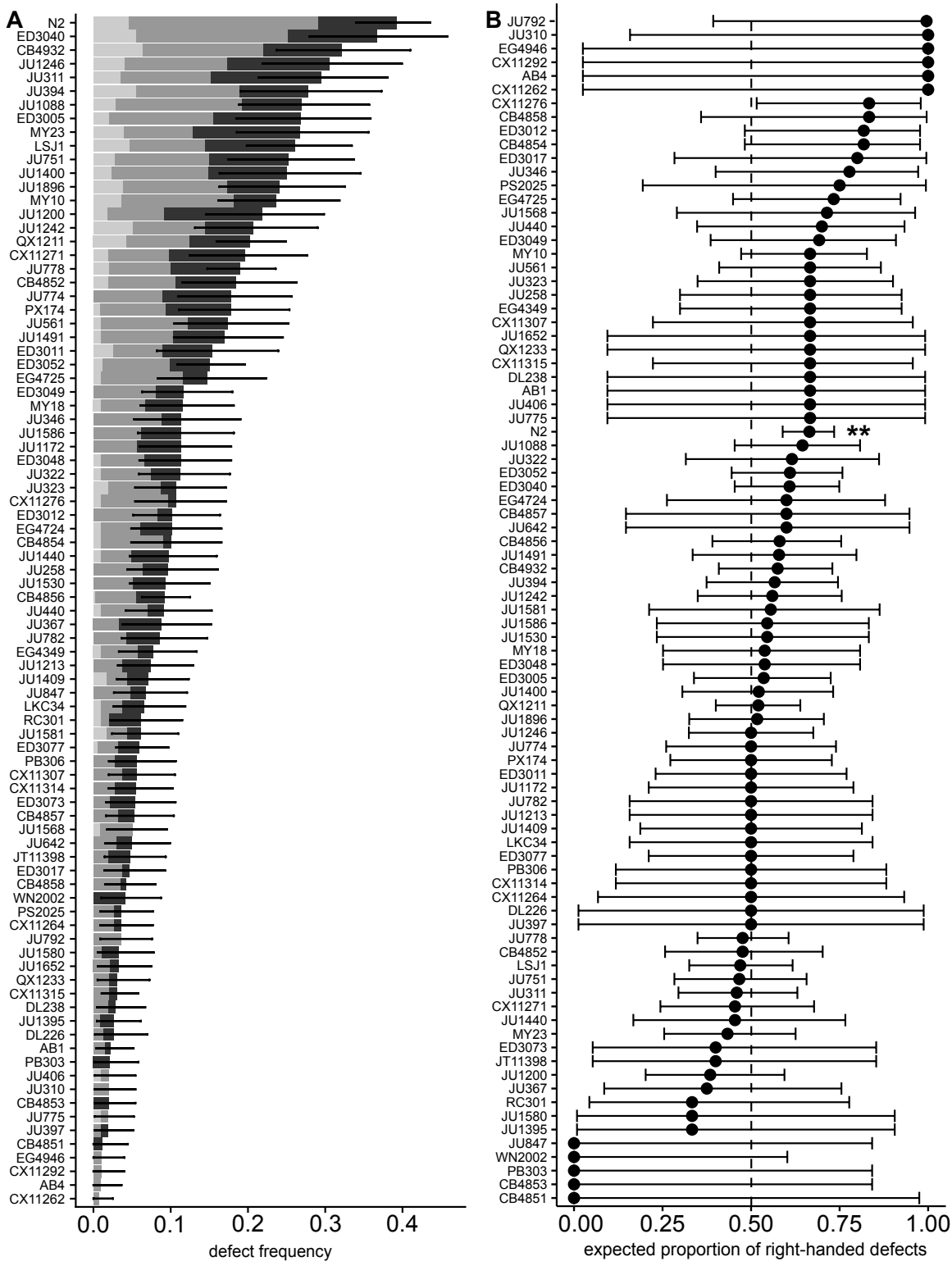

A
N2
ED3040
CB4932
JU1246
JU311
JU394
JU1088
ED3005
MY23
LSJ1
JU751
JU1400
JU1896
MY10
JU1200
JU1242
QX1211
CX11271
JU778
CB4852
JU774
PX174
JU561
JU1491
ED3011
ED3052
EG4725
ED3049
MY18
JU346
JU1586
JU1172
ED3048
JU322
JU323
CX11276
ED3012
EG4724
CB4854
JU1440
JU258
JU1530
CB4856
JU440
JU367
JU782
EG4349
JU1213
JU1409
JU847
LKC34
RC301
JU1581
ED3077
PB306
CX11307
CX11314
ED3073
CB4857
JU1568
JU642
JT11398
ED3017
CB4858
WN2002
PS2025
CX11264
JU792
JU1580
JU1652
QX1233
CX11315
DL238
JU1395
DL226
AB1
PB303
JU406
JU310
CB4853
JU775
JU397
CB4851
EG4946
CX11292
AB4
CX11262
0.0
0.1
0.2
0.3
0.4
defect frequency
B
JU792
JU310
EG4946
CX11292
AB4
CX11262
CX11276
CB4858
ED3012
CB4854
ED3017
JU346
PS2025
EG4725
JU1568
JU440
ED3049
MY10
JU561
JU323
JU258
EG4349
CX11307
JU1652
QX1233
CX11315
DL238
AB1
JU406
JU775
N2
**
JU1088
JU322
ED3052
ED3040
EG4724
CB4857
JU642
CB4856
JU1491
CB4932
JU394
JU1242
JU1581
JU1586
JU1530
MY18
ED3048
ED3005
JU1400
QX1211
JU1896
JU1246
JU774
PX174
ED3011
JU1172
JU782
JU1213
JU1409
LKC34
ED3077
PB306
CX11314
CX11264
DL226
JU397
JU778
CB4852
LSJ1
JU751
JU311
CX11271
JU1440
MY23
ED3073
JT11398
JU1200
JU367
RC301
JU1580
JU1395
JU847
WN2002
PB303
CB4853
CB4851
0.00
0.25
0.50
0.75
1.00
expected proportion of right-handed defects

Figure 3.2. *egl-1* RNAi reduces ray defects in some isotypes.
Overall propensity for defects given each condition varies in N2, JU311 and QX1211. *p<0.05, n.s. indicates not significant. Comparisons were calculated using Fisher's exact test. P-values were adjusted for multiple comparisons with 5% false-discovery rate.

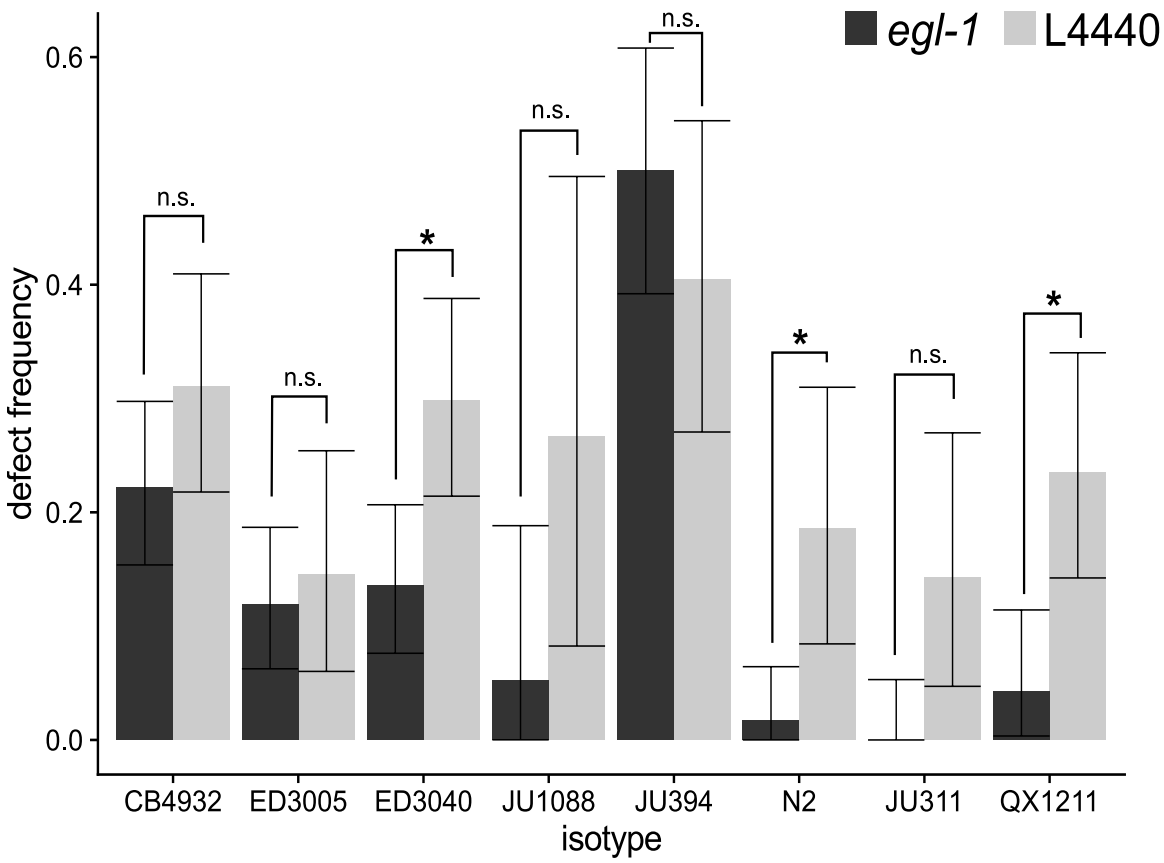

egl-1
L4440
defect frequency
0.0
0.2
0.4
0.6
n.s.
n.s.
*
n.s.
*
n.s.
*
n.s.
CB4932
ED3005
ED3040
JU1088
JU394
N2
JU311
QX1211
isotype

Figure 3.3. EMMA identifies 1 QTL on chromosome II.
X-axis indicates physical location (in Mb) and y-axis shows log10-transformed p-values. Black bar represents 5% genome-wide error rate using Bonferroni correction for multiple comparisons (= -log10 (0.05/4690) = 4.97×10^{-5}).

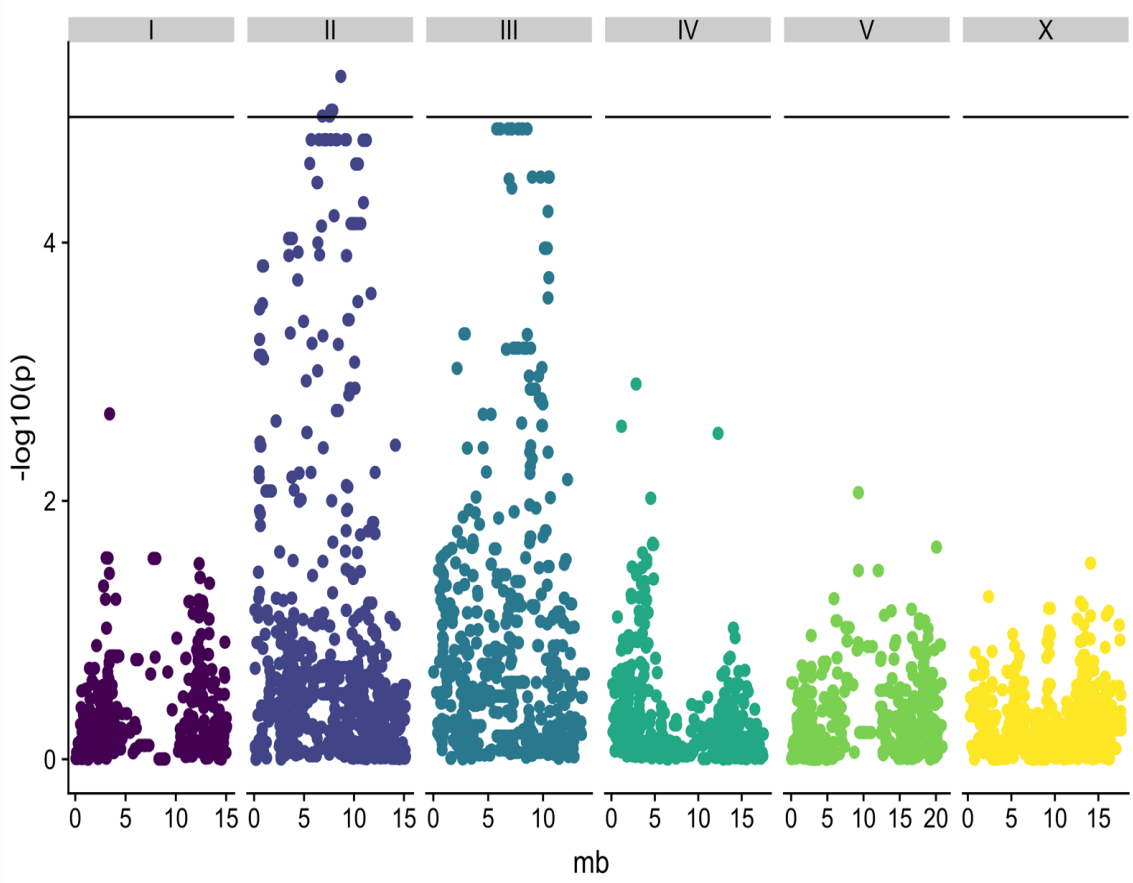

I
II
III
IV
V
X
-log10(p)
0
2
4
0 5 10 15
0 5 10 15
0 5 10
0 5 10 15
0 5 10 15 20
0 5 10 15
mb

Figure 3.4. Effect plots for each significant SNP identified by GWAS.
Y-axis shows variance for isotypes with either major or minor allele (X-axis). Isotypes with minor allele tend to significantly less defects at each SNP identified by GWAS using a mixed model.

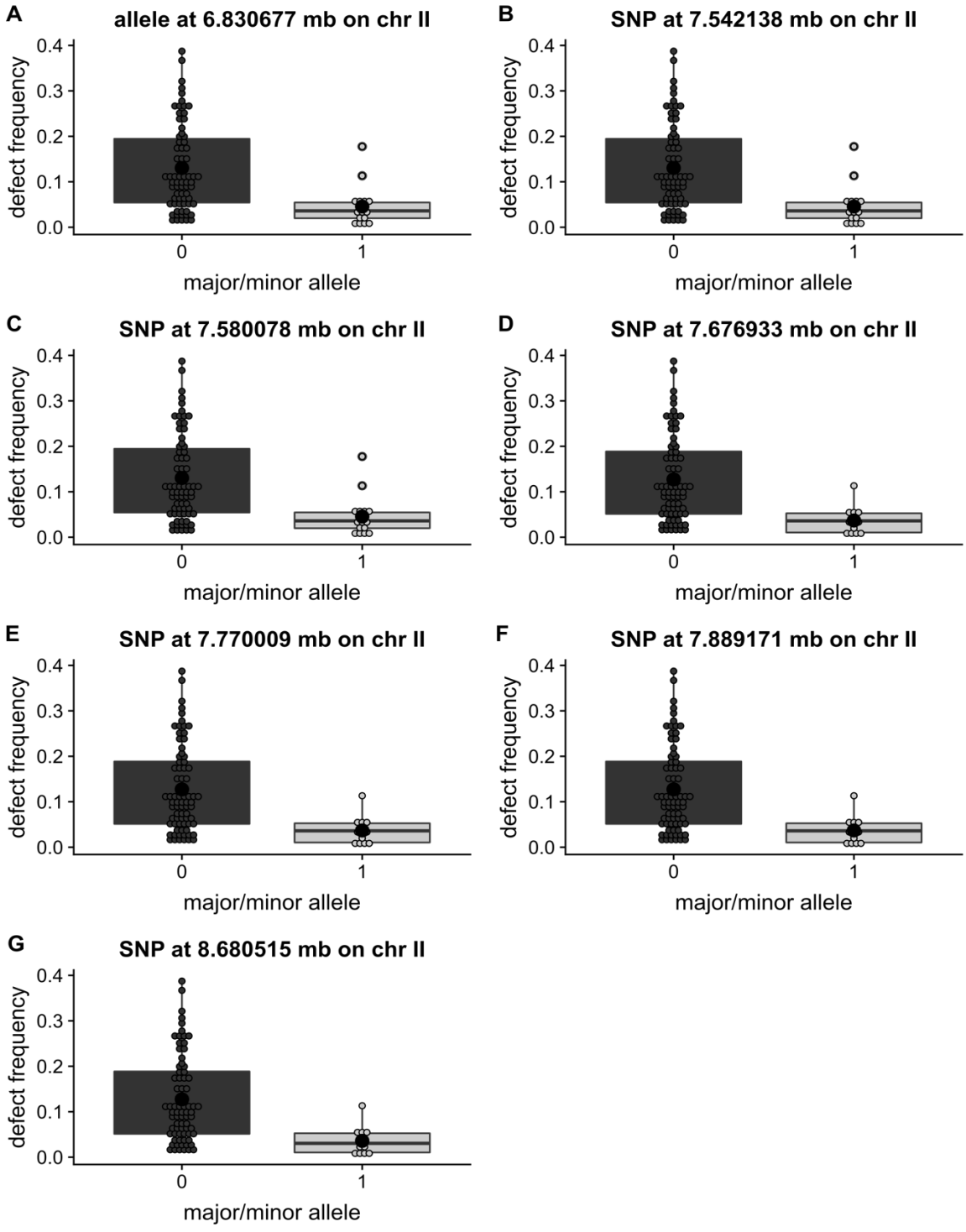

A
allele at 6.830677 mb on chr II
B
SNP at 7.542138 mb on chr II
C
SNP at 7.580078 mb on chr II
D
SNP at 7.676933 mb on chr II
E
SNP at 7.770009 mb on chr II
F
SNP at 7.889171 mb on chr II
G
SNP at 8.680515 mb on chr II
defect frequency
0.0
0.1
0.2
0.3
0.4
0
1
major/minor allele

Figure 3.5. Propensity for stochastic ray loss trends towards transgressive segregation.
Left, right bilateral defects were measured and shown are 95% Bayes binomial confidence intervals for the overall propensity for stochastic ray loss. $30 < n < 306$ per RIL, 1-3 samples per RIL or parental strain.

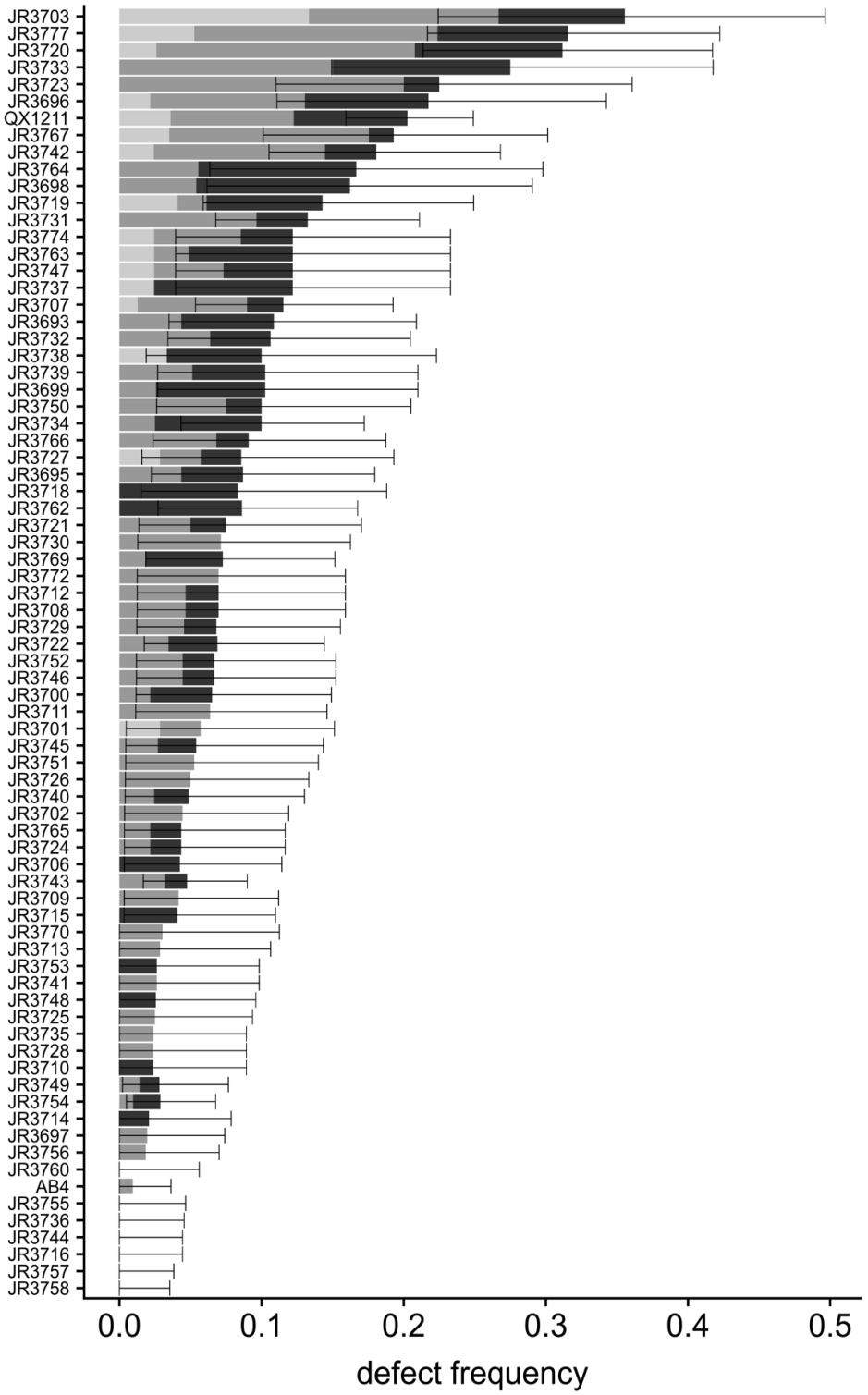

JR3703
JR3777
JR3720
JR3733
JR3723
JR3696
QX1211
JR3767
JR3742
JR3764
JR3698
JR3719
JR3731
JR3774
JR3763
JR3747
JR3737
JR3707
JR3693
JR3732
JR3738
JR3739
JR3699
JR3750
JR3734
JR3766
JR3727
JR3695
JR3718
JR3762
JR3721
JR3730
JR3769
JR3772
JR3712
JR3708
JR3729
JR3722
JR3752
JR3746
JR3700
JR3711
JR3701
JR3745
JR3751
JR3726
JR3740
JR3702
JR3765
JR3724
JR3706
JR3743
JR3709
JR3715
JR3770
JR3713
JR3753
JR3741
JR3748
JR3725
JR3735
JR3728
JR3710
JR3749
JR3754
JR3714
JR3697
JR3756
JR3760
AB4
JR3755
JR3736
JR3744
JR3716
JR3757
JR3758
0.0
0.1
0.2
0.3
0.4
0.5
defect frequency

Figure 3.6. Standard interval mapping identifies 3 independent and highly heritable QTL.

(A) Standard interval mapping using 1599 binary SNP markers indicates 3 QTL on chromosomes II, III and X are significant. Black bar represents 5% genome-wide significance threshold determined by N=10,000 permutations. (B) Effect plots at the most significant SNP (largest change in logarithmic-odds ratio (LOD) calculated in R/qtl on each chromosome (II, III and X)). RILs with alleles from parent AB4 have significantly less defects on II and III while the effect is reversed on chromosome X. (C) Model depicting 95% Bayes confidence intervals calculated for Standard Interval mapping (blue), Stepwise Haley-Knott regression in a multi-QTL model (yellow) and region identified by GWAS using linkage disequilibrium (green). The most significant SNPs identified by each method is represented by an asterisk in each respective color.

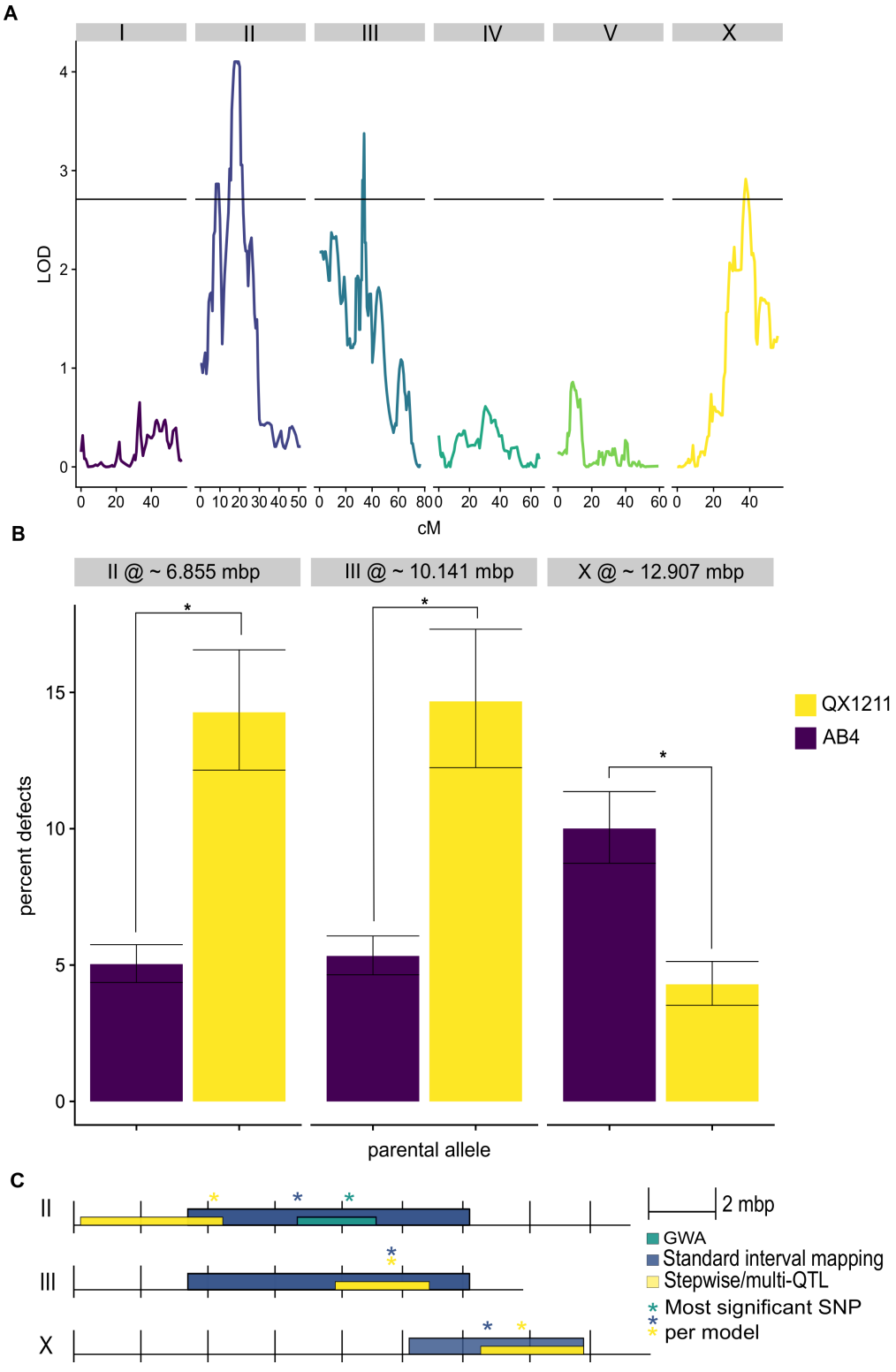

A
I
II
III
IV
V
X
LOD
cM
B
II @ ~ 6.855 mbp
III @ ~ 10.141 mbp
X @ ~ 12.907 mbp
percent defects
parental allele
QX1211
AB4
C
II
III
X
2 mbp
GWA
Standard interval mapping
Stepwise/multi-QTL
Most significant SNP
per model

Figure 3.7. Near isogenic lines (NILs) exclude a small region on chromosome II.

(A) Parental allele determined by Sanger sequencing at each marker is indicated by yellow (QX1211) or purple (AB4) boxes along the length of chromosome II and X. (B) Overall propensity of defects from F1 reciprocal crosses, near-isogenic lines on chromosome II and X and both parental strains. Black bars represent +/- 2 SEM. All genotypes were compared against each other using Fisher's exact test with 5% false discovery rate (C) 95% binomial confidence intervals from 1-sample binomial tests using resampling. F1 males from QX1211 males x AB4 hermaphrodites have significantly higher proportion of defects and potentially a right-hand bias like N2. *p<0.05,**p<0.01,***p<0.001, and lack of any marker indicates a non-significant comparison ($p > 0.05$).

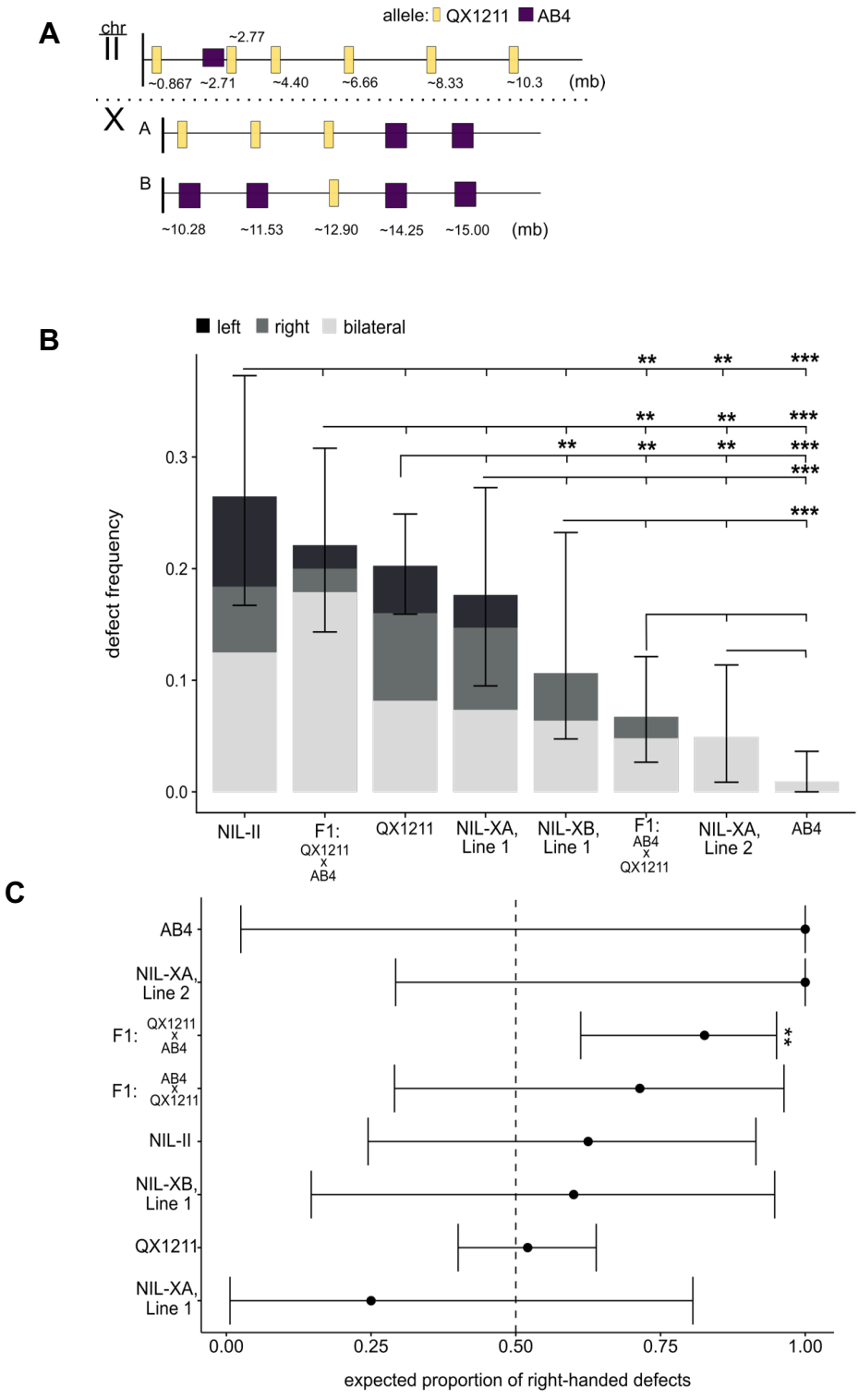
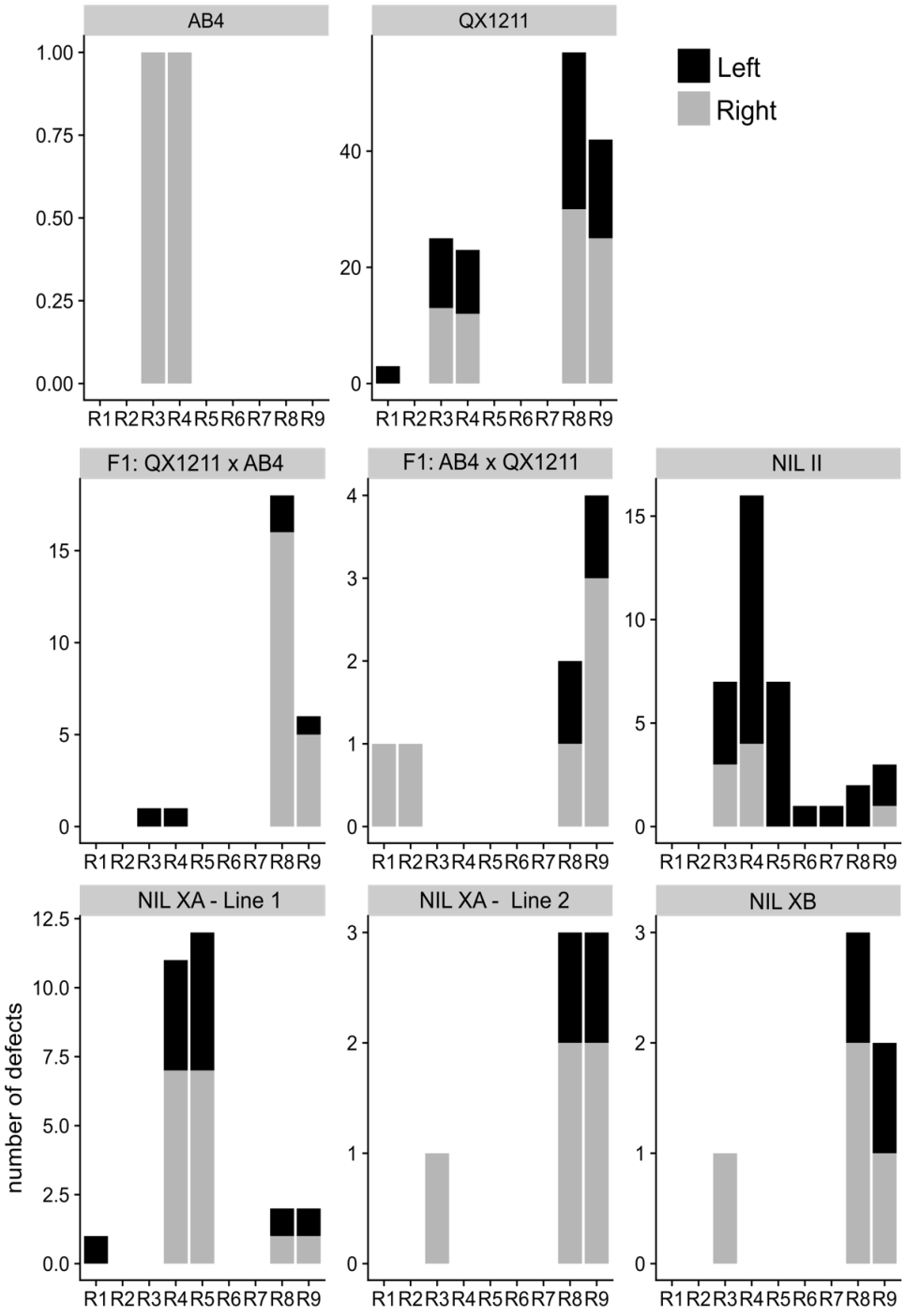

A
allele: QX1211 AB4
chr
II
~2.77
~0.867 ~2.71 ~4.40 ~6.66 ~8.33 ~10.3 (mb)
X
A
B
~10.28 ~11.53 ~12.90 ~14.25 ~15.00 (mb)
B
left right bilateral
defect frequency
0.0 0.1 0.2 0.3
NIL-II
F1: QX1211 x AB4
QX1211
NIL-XA, Line 1
NIL-XB, Line 1
F1: AB4 x QX1211
NIL-XA, Line 2
AB4
C
AB4
NIL-XA, Line 2
F1: QX1211 x AB4
F1: AB4 x QX1211
NIL-II
NIL-XB, Line 1
QX1211
NIL-XA, Line 1
0.00 0.25 0.50 0.75 1.00
expected proportion of right-handed defects

Figure 3.8. Specific rays affected by stochastic cell death varies in near isogenic lines.
The specific ray(s) affected by each defect were scored. Rays that appear fused together were counted as two separate defects.


AB4
QX1211
Left
Right
F1: QX1211 x AB4
F1: AB4 x QX1211
NIL II
NIL XA - Line 1
NIL XA - Line 2
NIL XB
number of defects
R1 R2 R3 R4 R5 R6 R7 R8 R9

Chapter Four

From ‘Heterotaxy in *Caenorhabditis*: widespread natural variation in left-right arrangement of the major organs’

Summary

While the arrangement of internal organs in most metazoans is profoundly left-right (L/R) asymmetric with a predominant handedness, rare individuals show full (mirror-symmetric) or partial (heterotaxy) reversals. The development of the nematode *C. elegans* is highly determinate, leading to a stereotyped L/R arrangement of the major internal organs (gut and gonad) in the laboratory strain. However, we found that the L/R orientation of the gut and gonad, but apparently not other major L/R asymmetries, varies in both sexes between natural isolates of the species. Whereas no reversals are observed in some isotypes, heterotaxy is present in up to ~11% of males and 12% of hermaphrodites in many others. In hermaphrodites, L/R reversals of the gut and gonad can involve one or both bilaterally asymmetric gonad arms. The observed heterotaxy in *C. elegans* males is unlikely to result from relaxed selection for male function in a hermaphroditic species, as we also observed this trait in both gonochoristic and hermaphroditic *Caenorhabditis* species. In many isolates, heterotaxy increases at elevated temperature. One strain shows a pre-gastrulation temperature-sensitive period for hermaphrodite reversals, suggesting a very early embryonic or germline effect on this much later developmental outcome. Genome-wide association studies of male heterotaxy across 97 isotypes showed that variation for this trait is associated with three genomic regions, and analysis of recombinant inbred lines from strains that span the spectrum of observed phenotypes suggest that a small number of loci are responsible for the variation observed. These findings reveal that heterotaxy is a widely varying quantitative trait in an animal with an

otherwise highly stereotyped anatomy, demonstrating unexpected plasticity in L/R arrangement of the major organs even in a simple animal.

Introduction

The arrangement of internal organs in most metazoans is substantially left–right (L/R) asymmetric with a largely invariant handedness. Molecular and cellular mechanisms that break L/R symmetry and subsequently determine internal organ placement have been illuminated in a variety of animal systems (1-7). While these mechanisms are largely conserved within a phylum, it is clear that the control systems that create and implement L/R asymmetry and handedness are widely divergent between phyla (159, 160) Although the innovation of handedness-determining systems appears to have occurred frequently during metazoan evolution, the evolutionary basis for establishing new mechanisms of handedness asymmetry is generally unknown. In the context of L/R axis formation, two major questions persist: how is a particular body plan selected for on evolutionary timescales (2)? And how is this L/R positional information established and maintained during development (84, 161–167)?

The asymmetry of the L/R axis is established during early embryogenesis and persists throughout development, as reflected in the overall body plan. However, subsequent symmetry breaks that are largely independent of earlier L/R asymmetries give rise to additional L/R differences, including in specific structures (e.g. handedness of heart looping (168, 169), laterality in the central nervous system (170–172) and motor handedness (118, 173–175). In humans, complete L/R reversal of internal organ

structure (*situs inversus totalis*) occurs in rare individuals, generally with no known ill consequences. However, disruption of internal organ asymmetries can result in birth defects with varying severity, most notably in cases in which two or more organs deviate from the canonical structure or position (176), a condition known as heterotaxy, which is frequently associated with ciliopathies (176, 177). The multiplicity of L/R symmetry breaks that generate anatomical, neuronal and behavioral bilateral asymmetries occurs even in simple animals, including the nematode *Caenorhabditis elegans* (167, 178–181), which contains only approximately 1000 somatic cells. The major structural chirality of the *C. elegans* embryo is established during the division of the two anterior granddaughters of the zygote (75). Anteroposterior skewing of mitotic spindles with a defined handedness, driven by actomyosin-directed chiral cortical flows that generate torque as these cells divide (76), creates an embryo with defined ('dextral') chirality. As a result, progeny cells subsequently make bilaterally asymmetric contacts with Notch-signaling cells, resulting in differential cell fate specification on opposite sides on the embryo (182). Many later L/R asymmetries, including differences in cell identities, specificity of bilateral pairs of neurons (183) and the stereotyped L/R arrangement of the gut and gonad (75) are prefigured by this early symmetry-breaking event. However, L/R differences between some cell types that are made stochastically , and the L/R motor handedness bias of male mating (118), are independent of this embryonic chirality. Further, we found that the stereotyped L/R arrangement of the intestine and gonad of males in the laboratory isolate is frequently reversed at elevated temperature, independent of the overall anatomical handedness of the animal (electronic supplementary material, table S1 and figure S2; (86)),

suggesting that the symmetry-breaking mechanism leading to the bilateral orientation of these organs is at least partially separable from the event that creates chirality of the embryo.

Identifying mechanisms that engender L/R asymmetries independent of embryonic chirality, and an understanding of the evolutionary basis for changes in these events, can be illuminated by analyzing potential incipient changes in the processes between evolutionarily divergent isolates of a species with variation in L/R asymmetries. We report here substantial variation in the arrangement of the major organs (the gut and gonad) in both sexes among a set of naturally inbred *C. elegans* wild isolates representing 97 distinct haplotypes, as well as in other *Caenorhabditis* species, including both hermaphroditic and gonochoristic species. Although all descendants from each of these naturally inbred *Caenorhabditis* isolates are isogenic, we found that heterotaxy occurs between genetically identical individuals of many isolates. Genome-wide association studies reveal that this frequent heterotaxy is associated with three chromosomal regions, demonstrating that heterotaxy is a genetically heritable trait and a substrate for natural selection. These findings reveal that heterotaxy of the two major organs is a common trait in an animal with an otherwise highly determinate anatomy, and that the L/R arrangement of these organs is subject to substantial variation during the radiation of a single species.

Results

The fidelity of left‐right gut/gonad asymmetry varies widely across wild isolates of *Caenorhabditis elegans* males

The chirality of *C. elegans* embryos is established during the four- to six-cell stage of embryogenesis, as a result of skewing of the mitotic spindles in descendants of the anterior daughter of the zygote, AB (75). This symmetry break, which occurs with a virtually invariant L/R handedness (‘dextral’ chirality), has been shown to lead to nearly all of the stereotyped L/R asymmetries in cell type and organ arrangement in the mature animal (50). This bilateral asymmetry includes the reproducible L/R orientation of the major internal organs, the gut and gonad, which is readily scored at low magnification in a dissecting microscope and has been used as a proxy for overall L/R anatomical handedness (85). Healthy animals of either sex with sinistral anatomical handedness can be generated as a result of physical manipulation of the mitotic spindles, perturbation of the integrity of the eggshell or exposure to low temperature (85). In addition, viable sinistral animals can arise as a result of mutations in *gpa-16* (84), which encodes a small rho-GTPase required for normal spindle placement, and in the actin cytoskeleton (179), leading to reversal of AB spindle skewing. However, several studies of a large number (more than 15,000) of animals (75, 84–86) reared at 20˚C indicated that such L/R (left – right) reversals in the arrangement of the major organs in the laboratory N2 isolate do not occur under normal conditions, suggesting that evolutionary constraints have led *C. elegans* to adopt a virtually invariant dextral body plan.

We previously reported that the L/R gut/gonad asymmetry in N2 males raised at 25˚C is frequently (more than 5%) reversed (86), with the fully developed gut occupying the position on the right normally taken by the gonad (Figure 4.1). These reversals do not appear to reflect a complete inversion of L/R anatomy, as a marker of the right specific ASER neuron is expressed in its normal position on the right (181). We also observed reversed males in the Hawaiian (Hw) isolate CB4856 and in recombinant inbred advanced intercross lines (RIAILs) derived from N2/Hw crosses (184, 185) under standard conditions, suggesting that this effect does not simply reflect breakdown of the L/R-determining system for these organs as a result of long-term culturing in the laboratory. These findings led us to ask whether such reversals might also occur naturally in the many available naturally inbred wild isolates of the *C. elegans* species, which show variation in a number of other traits, including vulva development (186, 187), susceptibility to pathogens and anti-parasitic resistance (41). Indeed, we found that among 100 wild isolates, there is widespread variation in the propensity for gut/gonad reversals in adult males: while 12 isolates never showed reversals (electronic supplementary material, table S1, n = 90 per isolate), most isolates showed significant rates of reversals with frequencies as high as 11.2% (JU778, n = 35/314, Figure 4.2, numerical data and confidence intervals shown in the electronic supplementary material, table S1). Thus, although *C. elegans* development is generally seen to be highly stereotyped, the L/R arrangement of the major organs in males is substantially variable in many wild isolates of the species (observed for a total of 474 out of 13,712 males).

Frequent reversals in male gut/gonad asymmetry occur in both hermaphroditic and gonochoristic *Caenorhabditis* species

We considered the possibility that natural variation in male L/R development of *C. elegans* isolates might be the result of reduced selective pressure on the male body plan in the species, which is hermaphroditic and can therefore effectively propagate in the absence of males. Consistent with such a possibility, we found that males of the hermaphroditic species *C. briggsae* also showed frequent L/R reversals (Figure 4.3). However, we also found that three of four closely related gonochoristic *Caenorhabditis* species (Figure 4.3) analyzed, *C. japonica*, *C. remanei* and *C. portoensis*, showed frequent gut/gonad reversals in males, even under conditions (20°C) in which N2 males never showed such reversals (Figure 4.2) Thus, the substantial variation in L/R orientation of the major organs is common in males of *Caenorhabditis* species, even when sex is essential for their propagation.

Extensive variation in L/R orientation of each gonad arm in hermaphrodites

Given our finding that males of both hermaphroditic and gonochoristic species exhibit L/R gut/gonad reversals, we asked whether similar variation in the relative position of the internal organs might also occur in hermaphrodites. In contrast to the single-arm male gonad, the hermaphrodite gonad consists of two arms arranged in opposite anterior–posterior orientation (Figures 4.1 and 4.4). Each arm terminates in a spermatheca, both of which connect to a common uterus. In both sexes, the four-cell gonad primordium is present mid-ventrally in the newly hatched L1 larvae; however,

the postembryonic lineages and development differ substantially in the two sexes (49, 50, 81). Given that natural reversals were never observed in a large number of hermaphrodites (more than 15 000) in the N2 isolate (75, 84, 85), it is possible that coordinating the migration of two gonad arms might require a more robust and high-fidelity developmental program than in males. However, although we found that several isolates never showed reversals in hermaphrodites (seven isolates, $n > 100$ each, greater than or equal to two replicates per isolate), hermaphrodites from some isolates showed frequent gut/gonad reversals, ranging from 0.8% ($n = 381$) in isolate JU1088 to 12% (n = 200) in MY16 at 20°C (Figure 4.4). Thus, natural reversal in the L/R orientation of the major organs is a characteristic of both sexes.

The gut/gonad reversals that we observed in hermaphrodites frequently involved only a single gonad arm. In normal N2 hermaphrodites, the anterior arm resides on the right and the posterior on the left, when viewed from the ventral aspect (Figure 4.3). We observed three classes of gut/gonad reversals in hermaphrodites: those involving only the anterior arm ('anterior heterotaxy'), only the posterior arm ('posterior heterotaxy') or both arms ('complete reversal'), in which the overall orientation of the gut and gonad is a mirror image of that in a normal dextral hermaphrodite (Figures 4.1 and 4.3). The frequency for each type of defect varied from 0% to 6% at 20°C ($n > 200$ for each isolate; Figure 4.4). Reversals involving only one arm (Figure 4.4) were substantially more prevalent than those involving both, suggesting that the event leading to these reversals may occur independently during gonad development

All classes of gut/gonad reversals in hermaphrodites of several of the isolates were significantly more frequent at 25°C than at 20°C (e.g. as was particularly evident for isolates JU1088, JU1530, and JU778; Fisher's exact test, $p < 0.05$, $p < 0.05$, $p < 0.001$ respectively, Figure 4.4). This might indicate inherent temperature sensitivity in the process of gonad migration, possibly akin to the temperature sensitivity of the nuclear migration of the ventral P epidermal cells during early larval development (188). However, some isolates showed no significant enhancement in the rate of L/R gut/gonad reversals (and in some cases, showed a lower rate of reversals at the higher temperature, albeit at insufficient sample size to provide statistical significance), suggesting a more complex and varied relationship between establishment of this L/R asymmetry and temperature.

Heterotaxy is independent of the event that establishes embryonic chirality and overall anatomical handedness

The overall anatomical L/R asymmetry is established in the early embryo, when the mitotic spindles of the daughters of the AB blastomere skew along the A/P axis with a dextral bias (75, 76, 167, 179). While this event leads to the defined dextral anatomical handedness of the animal, L/R anatomical handedness can be completely reversed (i.e. *situs inversus totalis*) by micromanipulation of the skewing spindles (75) and in *gpa-16* mutants, in which the orientation of the mitotic spindle is altered in the early embryo (84). To assess whether the reversals we observed reflect reversals in the chirality-inducing event in the early embryo, we took advantage of the reproducible

L/R asymmetry in the position of the coelomocytes, which can be scored in living animals by Nomarski microscopy. We analyzed ten isotypes for association of L/R reversals with overall anatomical handedness, using the ventral pair of coelomocytes ccAR and ccPR as an internal marker of the overall chirality of the animal (75). In hermaphrodites showing situs inversus totalis as a result of a temperature-sensitive mutation in *gpa-16(ts*), the anterior pair of coelomocytes (ccAR and ccPR) invariably (n = 10) are located on the left instead of at their normal position on the right, and the gut /gonad arrangement is sinistral (Figure 4.3). In contrast, for all 10 strains examined, the coelomocyte pair was invariably located in the normal right-side position, including in all animals with partial or complete gut/gonad reversals (n > 43 animals with reversals; Figure 4.3). These observations strongly suggest that reversal in the position of the gut and gonad occurs independently of the overall anatomical chirality established in the early embryo and hence reflects bona fide heterotaxy of the major organs.

An event preceding mid-embryogenesis can influence heterotaxy much later in development

Our observation that the propensity for L/R reversals in gut and gonad arrangement is influenced by temperature led us to evaluate the developmental stage at which the sensitivity to temperature for this effect occurs in hermaphrodites of isolate JU778, which shows among the highest rates of reversals. Synchronized populations of JU778 animals were grown at either 20°C or 25°C for at least two generations and

early pre-gastrulation embryos, present within the uterus of hermaphrodite adults, were temperature-shifted up or down and scored for heterotaxy (Figure 4.5). JU778 hermaphrodites maintained at 20°C or those upshifted during gastrulation to 25°C (Figure 4.5) showed similarly lower rates of heterotaxy (4.5%, n = 672 and 4.7%, n = 300, respectively; Fisher's exact test, 5% false discovery rate (128), q = 0.87). In contrast, animals grown continuously at 25°C (13.7% reversals, n = 475), and pre-gastrulation embryos that were downshifted from 25°C to 20°C (15.54%; n = 251) showed similarly (Fisher's Exact Test, 5% false discovery rate (128), q = 0.61) higher rates of heterotaxy (Figure 4.5) compared with those grown at 20°C (q < 0.0001, Fisher's exact test). This finding suggests that the cellular events leading to L/R reversals in response to temperature occur during early embryogenesis or perhaps in the maternal germline, long before the migration of the DTCs that occurs during the larval stages.

Genome-Wide Association Study (GWAS) identifies three major regions associated with L/R gut/gonad reversals

In an effort to characterize the genetic complexity leading to heterotaxy in males, we performed a genome-wide association study (GWAS) using linear mixed-model analysis, including a pairwise identity-by-state (IBS) kinship matrix to correct for population structure (143) (based on the distribution of this phenotype across 100 wild isolates (Figures 4.2 and 4.6). Owing to the highly variable nature of this phenotype (Figure 4.2, electronic supplementary material figure S1 and table S1 from manuscript

(87)), we initially compared the frequency of reversals for all 100 natural isolates against each other before association mapping. The majority of significant comparisons were found between isolates at the extreme ends of the phenotypic spectrum, where the only significant comparisons were between isolates with 0% and approximately 8–11% L/R reversals (Fisher's exact test, 5% false discovery rate, $q \leq 0.049$). In order to identify any loci associated with the propensity for heterotaxy, we chose to treat the phenotype as a binary, categorical variable and used this category for association mapping. Thus, to perform the GWAS, we collected the 12 isolates with reversal frequencies of 0% into one group (electronic supplementary material, table S1 from original manuscript (87)) and the remaining 88 isolates into another.

By applying this binary/categorical phenotype, we found that GWAS identified one highly significant major region on chromosome II (23 SNPs) and two significant regions on chromosome III (nine SNPs total). The region of chromosome II spans 10.75 Mb (0.55 Mb to approx. 11.3 Mb), and all SNPs significantly associated with the phenotype in this region were found to be in near-perfect linkage disequilibrium (LD; $R^2 = 0.88$) (189). The first peak on chromosome III spans a 4.3 Mb region (2.61–6.91 Mb) and all significantly associated SNPs also show moderate LD ($R^2 = 0.58$). The second peak on chromosome III spans a small, 0.4 Mb region (10.1 Mb to approx. 10.5 Mb), which is also in nearly perfect LD ($R^2 = 0.75$), but only in moderate LD with the other region on chromosome III ($R^2 = 0.53$). Although the region on chromosome II and first region on chromosome III have sufficiently large recombination intervals

(19.1 cM on chromosome II, first region on chromosome III approx. 10.29 cM) that there would, in principle, be expected to be frequent recombination events separating the causal loci from linked SNPs, it has been shown that *C. elegans* natural isolates have recently undergone selective genomic sweeps and hence share large genomic segments (145). Thus, rare variants affecting the variability of heterotaxy that have accumulated in these regions may be inherited as large shared blocks that include the causal loci.

We sought to estimate heritability of this trait (190–193) by analyzing the genetic and random (owing to measurement error) variance for each of 4690 SNPs using restricted maximum-likelihood method (REML). We included an IBS kinship matrix in the REML estimates to control for population structure. This analysis yielded a heritability (h^2) of 0.302. However, this estimate may be slightly biased and/or artificially inflated: as interaction and dominance terms cannot be parsed from GWAS alone, our analysis considers only additive effects for each SNP. Moreover, heterotaxy is significantly affected by temperature with many isolates (Figure 4.4 and electronic supplementary material, figure S1 (87)), indicating an environmental component or genotype–environment interaction that is not included in our heritability estimate (Figure 4.6).

The major variation in heterotaxy between a strain with no reversals and one with frequent reversals is attributable to a small number of loci

The GWAS analysis suggested that there are at least three loci that influence the prevalence of heterotaxy across the 100 wild isolates. However, GWAS identifies only those genomic regions that vary in a substantial number of isolates. We sought to assess the genetic complexity of this trait further by analyzing its segregation following a cross between an isolate with no observed heterotaxy (AB4) and one that shows a high incidence for the trait (QX1211). Recombinant inbred lines (RILs) (141, 194) were generated from this cross and the frequency of L/R gut/gonad reversals in males was analyzed (Figure 4.7 and electronic supplementary material, figure S4 and table S2 (87)) Of 72 RILs, 29 (approx. 40%) showed no reversals, suggesting relatively low genetic complexity for the trait. The remaining RILs showed reversal frequencies of up to 10%, with no strong evidence of transgressive segregation (Fisher's exact test, 5% false discovery rate (128), $q > 0.05$). These results indicate that heterotaxy is a genetically tractable trait for which there may be one or a small number of loci underlying the phenotype in the QX1211 isolate; however, further analysis will be required to determine the effect of each locus *in vivo*.

Discussion

We report five major advances that reveal both the complexity of events leading to L/R asymmetry in the arrangement of organs in a simple metazoan, and the evolutionary plasticity underlying it. (i) We found that, across a single species with a generally highly determinate development, the stereotyped arrangement of the two

major organs, the gut and gonad, is subject to wide variation in fidelity across wild isolates in both sexes: while some naturally inbred isolates never reverse the canonical arrangement, other isolates show frequent L/R reversals, without reversing the overall chirality of the animal established during early embryogenesis. Thus, these reversals reflect *bona fide* heterotaxy. (ii) L/R reversals can affect either the anterior or posterior arm of the hermaphrodite gonad, suggesting that the symmetry-breaking mechanism that affects the arrangement of these organs may be assessed independently for each domain of the organ. (iii) Frequent reversals are also observed in males of other nematode species, including those in which males are obligate for reproduction, implying that a tendency towards heterotaxy is not the result of relaxed selection of male function in hermaphroditic species. (iv) The propensity for heterotaxy varies with temperature and the temperature-sensitive period for this effect in one isolate occurs unexpectedly early, well before the stage at which the reversals are observable. Thus, an event in the very early embryo or maternal germline influences a symmetry break that is evident only much later in development. (v) Variation in the propensity for heterotaxy across the 100 isolates is attributable, in part, to differences in three regions of the genome, and a small number of genetic changes appear to account for the variation between one isolate with high rates of heterotaxy, and one that shows none. Thus, the variation in the frequency of heterotaxy in *C. elegans* may be a relatively simple genetic trait.

Our finding that heterotaxy is common in many wild isolates suggests that, although the overall embryonic chirality and a number of L/R asymmetries in the animal appear

to be obligately rigidly determinate, the specific L/R orientation of the two largest internal organs is not. Such naturally occurring heterotaxy in *C. elegans*, exceeding a rate of 10% in some natural isolates, is substantially greater than the prevalence of heterotaxy for the human population, which has been estimated to be one in 15,000 (15). However, it is interesting to note that, as we have observed with different isolates of *C. elegans*, the prevalence of heterotaxy varies substantially between different groups of humans (176) with a higher prevalence among individuals of Asian descent. Thus, the developmental fidelity of L/R organ packing is subject to substantial variation across different populations of these phylogenetically widely separated species.

In contrast with males, in which the gut/gonad orientation is either normal or fully reversed, reversals of the L/R arrangement of the anterior and posterior hermaphrodite gonad arms appear to be independently determined. It is not clear whether the uncorrelated heterotaxy between the two arms reflects an L/R symmetry-breaking process that occurs independently for both arms, or a single system that influences the probability that each arm will occupy its normal or reversed orientation. In contrast, there appears to be some correlation between the prevalence of heterotaxy in males and hermaphrodites in the most strongly affected isolates. This correlation is not absolute, however, as some isolates that showed substantial heterotaxy in hermaphrodites (JU360, JU1088, ED3012) did not show male heterotaxy at 20°C. It will be of interest to assess the role of sex-determining mechanisms in influencing the L/R handedness-determining mechanisms of the major organs.

We found that the frequency of organ reversals is temperature-dependent in many of the wild isolates, indicating that there is an environmental component to the observed heterotaxy. Temperature-induced stress is widely known to lead to errors in development ((1) and reviewed in (195)) could explain the tendency towards increased heterotaxy in these isolates. However, this cannot fully account for the observed effect of temperature, as some isolates showed an indication of reduced heterotaxy at elevated temperatures. While the L/R heterotaxy we have observed in both sexes of *C. elegans* is evident well after embryogenesis, we were surprised to find that the sensitive period for the temperature-dependent event that influences the frequency of heterotaxy is during very early embryogenesis or even in the maternal germline. Thus, an event preceding organogenesis or even any overt differentiation affects a symmetry-breaking process that occurs only after the gut and gonad organs are formed and undergo maturation during larval development. It is conceivable that a chiral structure in the oocyte or early embryo, perhaps an element of the cytoskeleton, might establish an L/R asymmetry that is interpreted much later at the time that the distal end of the gonad begins to migrate, leading to elongation of the gonad arm.

A recurring theme emerging from recent studies on *C. elegans*, Xenopus and gastropods is that establishment of L/R differences in gene expression, and overall handedness, depends on the regulation of cytoskeletal rearrangements across these divergent organisms (196). However, it is unclear whether the particular pathways that regulate the cytoskeleton to direct L/R symmetry breaks are phylum specific. In *C. elegans*, the actomyosin cortex plays a critical role in establishing the overall L/R

anatomical asymmetry (75, 84, 167, 181) by generating chiral counter-rotating cortical flows along division axes, thereby generating torque that causes the spindles to skew clockwise during formation of the six-cell embryo (Figure 4.8A). This initial bilaterally asymmetric arrangement of the blastomeres establishes bilateral differences in lineages, fates and position of cells that persist into adulthood (75, 181, 197).

Might cytoskeletal rearrangements also influence symmetry-breaking events that occur later in development, including those that underlie the type of heterotaxy we have described here? Recently, Davison et al. (198) reported that the L/R axis in Xenopus and snails is most probably determined by the regulation of cytoskeletal elements that are predetermined and inherent to the chirality of the embryo itself (179, 199, 200). Further, it was reported that heterotaxy in Xenopus and chirality of snails could be caused by ectopically expressing formin, which regulates polymerization of the actin cytoskeleton. It may be noted that of at least seven *C. elegans* genes that encode proteins with formin-like domains, four (*exc-6*, *frl-1*, *fozi*-1 and *cyk-1* (reside on chromosome III, on which we have identified two significant regions associated with heterotaxy) (201–205). However, our initial genetic analysis is of insufficient resolution to point to specific causal mediators of gut/gonad heterotaxy. Further studies, using information from the RILs, will make it possible to delineate quantitative trait loci (QTL) that underlie heterotaxy in both sexes. It will be of interest to learn whether the causal genetic variations leading to heterotaxy in *C. elegans* play conserved roles in L/R organ arrangement in other Bilateria and whether there is any relationship between them and the L/R randomization of internal organs seen in humans, for example in Kartagener's syndrome (177).

Although handedness asymmetry in the arrangement of the visceral organs is particularly striking, a number of other L/R asymmetries in vertebrates, including brain laterality, ocular dominance and motor handedness (120, 121, 124, 206), are established by mechanisms that are largely (though not fully) independent during development. The findings reported here lend further support to the view that a multiplicity of independent symmetry-breaking events lead to distinct L/R asymmetries even in a very simple animal with less than 1000 somatic cells (summarized Figure 4.8). The first and major symmetry break, occurring at the 4–6 cell stage of embryogenesis, sets the stage for many subsequent L/R differences in cell positions and fates throughout the rest of development. However, a number of later symmetry-breaks that are independently assessed occur with particular cell types and behaviors. Stochastic establishment of asymmetric cell identities in the nervous system resulting from cellular interactions that are independent of the early chirality-establishing events occur during specification of the bilateral pair of AWC neurons: i.e. asymmetric calcium signaling mediated by innexins, claudins (78, 207) and the transcription factor DIE-1, establishes asymmetric expression of chemoreceptors in AWC neurons (207, 208) (Figure 4.8B). Further, *C. elegans* males exhibit a right- handed motor behavior bias during mating that is also independent of anatomical L/R asymmetry (118)(Figure 4.8C), implying the existence of yet another L/R symmetry-breaking process that affects motor output. The frequent heterotaxy we have observed suggests that there may exist yet another symmetry-breaking mechanism that affects the L/R arrangement of the major organs in the animal independent of embryonic chirality (Figure 4.5). That several, apparently independent, processes function to direct L/R

differences throughout development in this simple animal underscores the critical importance of distinguishing differences in bilateral structures and activities in many contexts during the formation of bilaterian metazoans.

Methods

General maintenance and worm strains used

Except where noted *C. elegans* isotypes and *Caenorhabditis* species were maintained as described by S. Brenner (125) and scored at room temperature. The strains corresponding to different *Caenorhabditis* species used in this study are N2 - *C. elegans*, CB5161 - *C. brenneri*, JU724 - *C. remanei*, DF5081 - *C. Japonica*, EG4788 - Sp.6, AF16 - *C. briggsae*. For a list of *C. elegans* isotypes used in this study, see NAS drive (87).

Male Stocks

Briefly, male stocks were generated by aerating ~ 40 L4-stage hermaphrodites in 7% ethanol (400µl) for 45 minutes and maintained on standard conditions (126). F1 males were then mated to sibling hermaphrodites and their progeny was used for data collection.

Scoring gut-gonad orientation

The gut-gonad orientation in males was performed as described previously (86). Briefly, 20-30 adult worms were anesthetized in 3µL of 5mM Levamisole on a 5% agarose pad. An eye-lash tool was used to orient males ventral-side up and slides

were then sealed. Mounted worms were imaged at 10x and 63x magnification on a Zeiss Axioskop 2. Hermaphrodite gut-gonad orientation was scored on unseeded NGM plates by allowing L4 or young adults to chemotax towards a spot of OP50 *E.coli* mixed with 1 M sodium azide for 1 hour at room temperature. Anesthetized worms were then rolled using an eye-lash tool and scored ventral side up on a Motic dissecting scope and imaged at 45x.

Temperature-shift experiments

Continuously fed stocks of JU778 were grown at 20˚C or 25˚C for 7 days (~2-3 generations) before temperature shifts. Synchronized L1 larvae were collected by hatching embryos from gravid adults treated with alkaline hypochlorite solution in M9 buffer and then grown on NGM plates seeded with OP50 bacteria at 20˚C or 25˚C until they were gravid adults but before any eggs were laid on the plates (151). They were then lysed using alkaline hypochlorite solution and the surviving pre-gastrulation embryos were collected. Embryos were then placed on NGM plates with food and allowed to develop at 20˚C or 25˚C. Gut-gonad orientation was then scored as described above.

Statistics

Data analyses were performed with R Statistical Software v. 3.2.3 (209).Fisher's Exact Tests were used to compare frequencies between isotypes unless otherwise stated. The False Discovery Rate method developed by Benjamini and Hochberg (128) was used to correct for multiple comparisons for all tests excluding GWAS.

Genome wide association (GWA)

A GWAS was performed using the EMMA package for R (153). Genotype information used was a binary SNP map with 4,690 markers spaced by approximately 100 kb per marker (145). Likelihood ratio tests were used to determine p-value per SNP using the emma.ML.lrt function. Mixed model associative tests also included an IBS kinship-matrix calculated using the binary SNP map, which accounts for relatedness and population structure in model organisms (39). The Bonferroni Method was used to account for multiple comparisons. The black line spanning all chromosomes in Figure 4.7 represents this cut-off at alpha = 0.05/4960 = 4.97×10^{-5}. The phenotypic input for mixed modeling was binary: isotypes with 0% reversals were assigned to an arbitrary group 0, and all other isotypes were assigned to an arbitrary group 1. 10,000 permutations of p-values were obtained from GWAS without accounting for population structure, increasing any potential noise in order to test significance of SNPs (Supplemental Figure 1 from (87)). Linkage disequilibrium for two loci identified by EMMA was calculated using PLINK software version 1.07 (210).

Figure 4.1. Schematic of gut/gonad heterotaxy in Caenorhabditis males and hermaphrodites.

The anterior pair of coelomocytes, ccAR and ccPR, (marked in red) is found on the right side of normal (dextral) worms and is used as reference for the L/R handedness. The black texture marks the gut, with the gonad is in white. All animals are shown in a ventral view with anterior up. (A) Dextral (normal) male: the coelomocytes and gonad are positioned on the right side of the animal, with the gut on the left. (B) Reversed male: the position of the gut and gonad are reversed with the coelomocytes at their normal right-side position.. (C) Dextral hermaphrodite: the anterior gonad arm is on the right side of the animal with the posterior arm on the left. (D) Complete gut/gonad heterotaxy: the positions of the anterior and posterior gonad arms and gut are L/R-reversed, while the anterior pair of coelomocytes remains on the right. (E) Posterior heterotaxy: the position of only the posterior gut and gonad is reversed compared to that in dextral hermaphrodites. (F) Anterior heterotaxy: the position of only the anterior gut and gonad is reversed compared to that in dextral hermaphrodites. Images drawn by ALS.

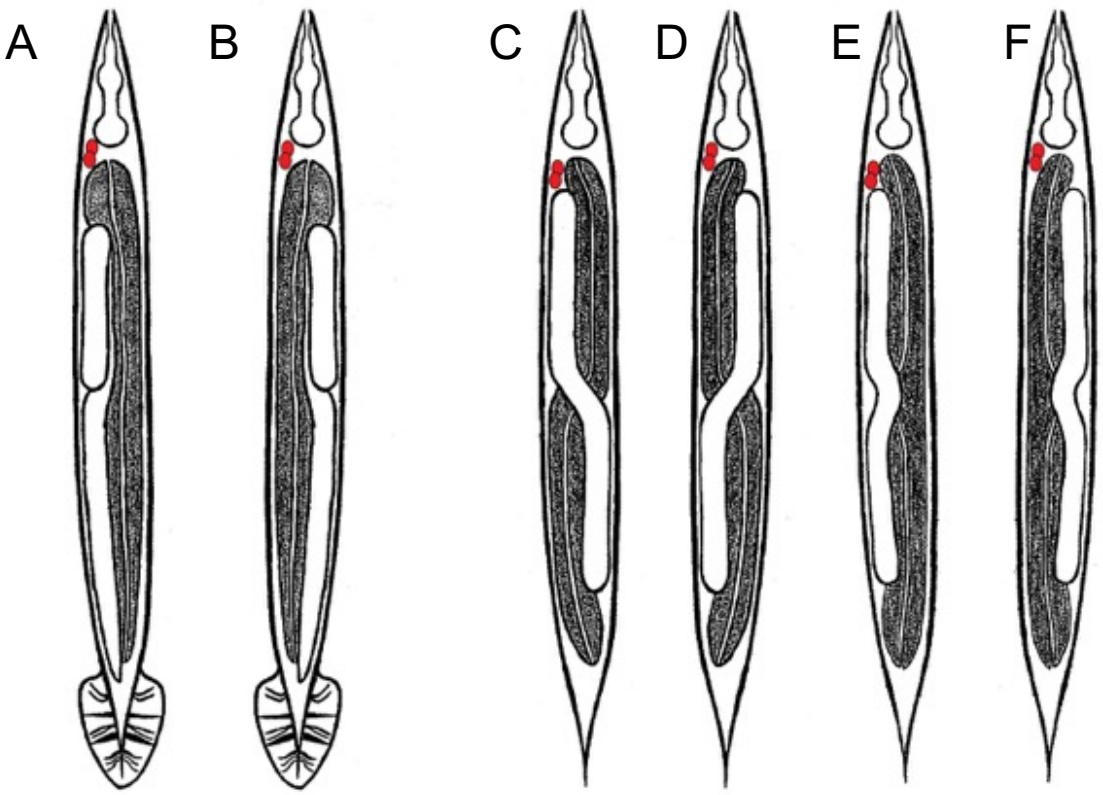

A
B
C
D
E
F

Figure 4.2. The propensity for L/R gut-gonad reversals varies widely in males of 100 *C. elegans* isotypes.
(A - C) Examples of DIC images of ventral view showing the relative position of the gonad. Coelomocytes are marked with white-dashed lines and the top of the gonad-arm is marked with black-dashed lines. Dotted red box in the adjacent cartoon denotes the corresponding region in the photomicrograph in reference to the rest of the body. (A) dextral N2 male. (B) dextral male from isotype JU778. (C) reversed JU778 male. (D) Percent of L/R gut/gonad reversals in adult males in the different natural isotypes grown at 20˚C. Each isotype is numbered 1-100 in rank order based on reversal frequency observed. Sample size per isotype ranges from 90 – 700 animals. See Supplemental Table 1 from original manuscript for the key to the numbered isotype strain list and statistical analysis (87). Data collected by DCC, BB and MRA.

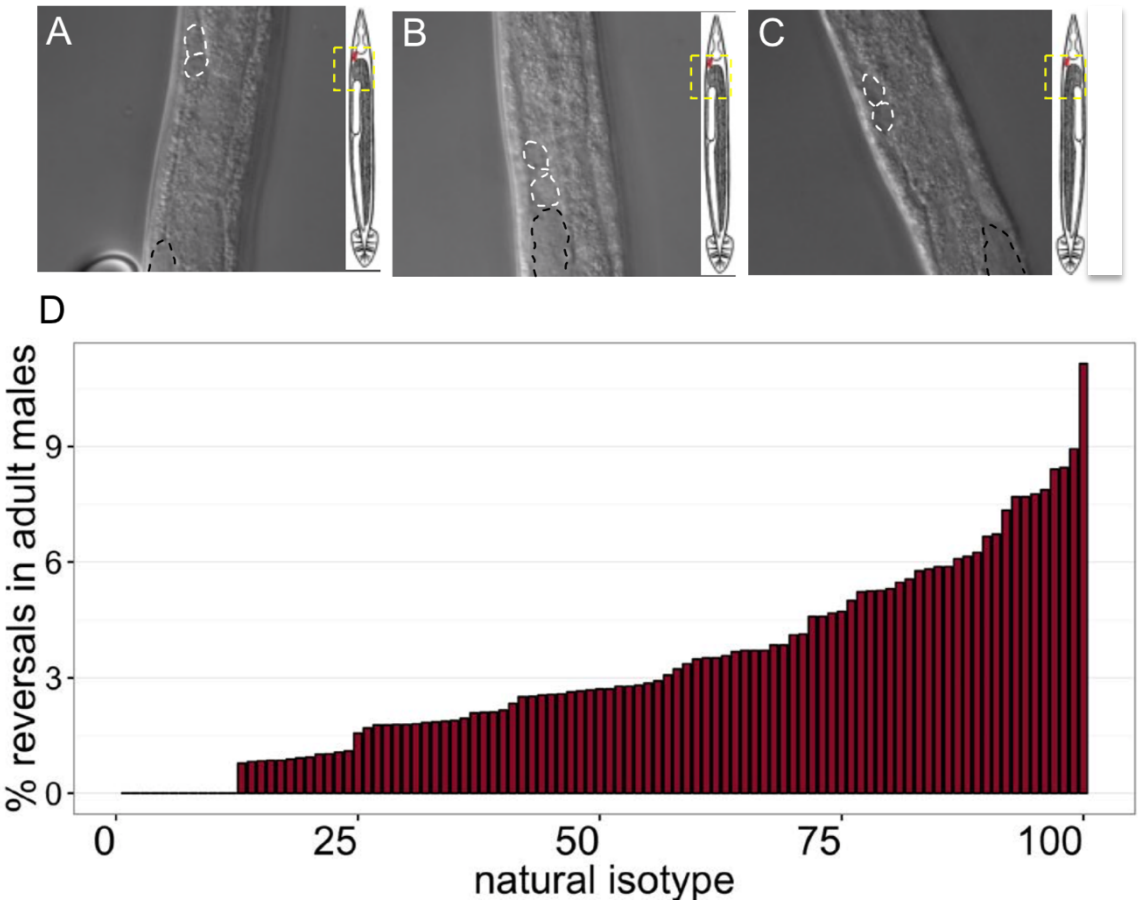

A
B
C
D
% reversals in adult males
0
3
6
9
0
25
50
75
100
natural isotype

Figure 4.3. Other Caenorhabditis species show gut/gonad reversals in males. (A) L/R Gut-gonad reversals in males of the indicated species. Black bars represent 95% confidence intervals calculated using the Clopper-Pearson method. Sample size is noted above each confidence-interval. Reproduction mode (gonochoristic or hermaphroditic) is indicated below each bar. (B) Abbreviated phylogenetic tree showing relationship of select *Caenorhabditis* species based on (112). Sp.6 has is referred to *C. portoensis* in Chapter 2. Data from DCC, figure made by YTC and MRA.

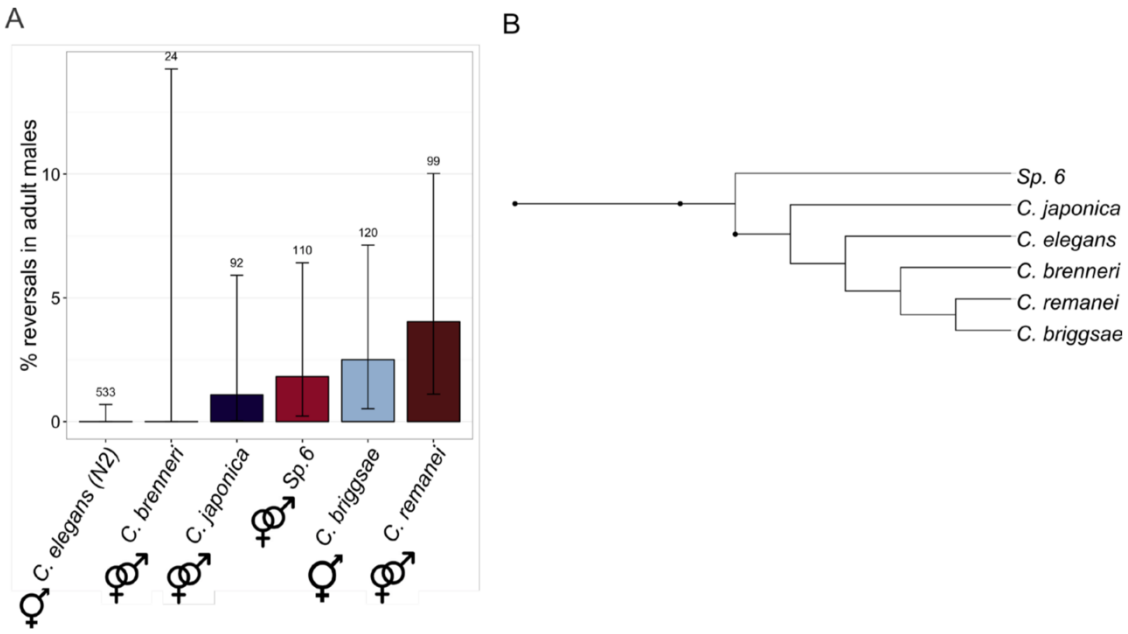

A
% reversals in adult males
0
5
10
533
24
92
110
120
99
C. elegans (N2)
C. brenneri
C. japonica
Sp. 6
C. briggsae
C. remanei
B
Sp. 6
C. japonica
C. elegans
C. brenneri
C. remanei
C. briggsae

Figure 4.4. Temperature effect on hermaphrodite heterotaxy.
(A-D) Ventral view of hermaphrodites (anterior up). White arrows denote position of gonad arms, which are visible as clear patches under the dissecting scope. Adjacent cartoons denote the type of heterotaxy observed, as described in Figure 4.1. A) Dextral hermaphrodite from isotype JU778. (B) Reversed hermaphrodite from isotype JU778. (C) Posterior heterotaxy from isotype JU778. (D) Anterior heterotaxy from isotype JU778 E-H: higher magnification DIC images showing the anterior coelomocytes ccAR and ccPR (white-dashed lines) and anterior portion of gonad-arm (black-dashed lines) (N2). (E) Dextral hermaphrodite from isotype N2. (F) gpa-16 (it143) hermaphrodite with fully sinistral handedness (situs inversus totalis). Note that both the anterior gonad-arm and the coelomocyte pair ccAR and ccPR are on the left side. (G) Dextral hermaphrodite from isotype JU778. (H) Complete heterotaxy in hermaphrodite from isotype JU778. The coelomocyte pair ccAR and ccPR remain on the right side of the animal, but the gut and gonad positions are reversed. (I) Frequency of each class of heterotaxy in hermaphrodites from a subset of *C. elegans* isotypes cultured at 20˚C and 25˚C. Black-lines represent 95% confidence intervals calculated using Clopper-Pearson/Exact Method. Sample size is noted above each condition. Data and images collected by ALS; figure made by MRA.

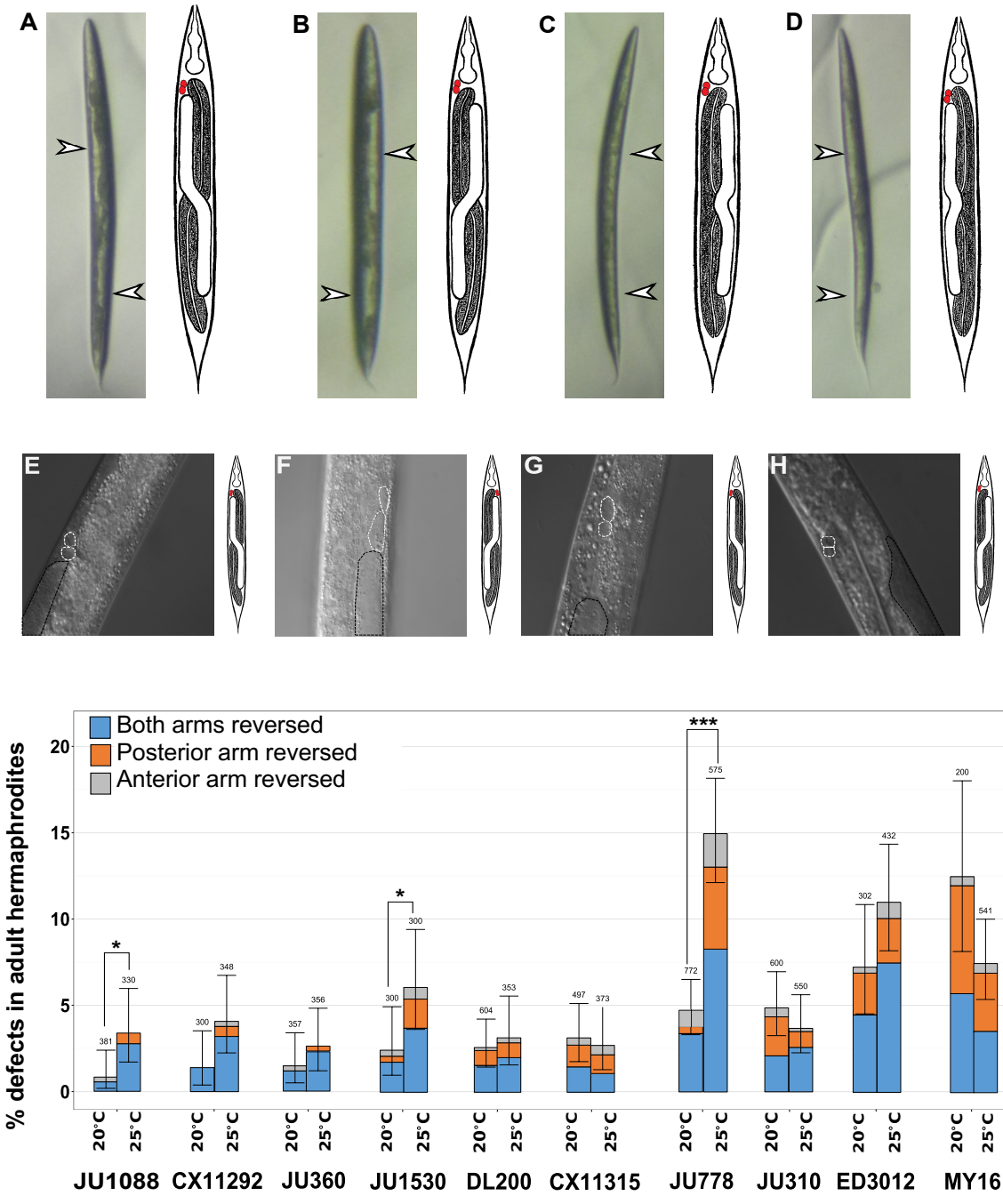

A
B
C
D
E
F
G
H
Both arms reversed
Posterior arm reversed
Anterior arm reversed
% defects in adult hermaphrodites
0
5
10
15
20
*
*
***
381
330
300
348
357
356
300
300
604
353
497
373
772
575
600
550
302
432
200
541
20°C
25°C
JU1088
CX11292
JU360
JU1530
DL200
CX11315
JU778
JU310
ED3012
MY16

Figure 4.5. Temperature sensitive period influencing heterotaxy coincides with in utero embryonic development.
Adjacent pie-charts show proportion of hermaphrodites with any heterotaxy and stacked bar shows proportion of each class of heterotaxy. (A) Pre-gastrulation-stage embryos cultured continuously to adulthood at 20˚C (n = 672). (B) Pre-gastrulation-stage embryos upshifted from 20˚C to 25˚C and cultured to adulthood at 25˚C (n = 300). (C) Pre-gastrulation-stage embryos cultured continuously to adulthood at 25˚C (n = 475). (D) Pre-gastrulation-stage embryos downshifted from 25˚C to 20˚C and cultured to adulthood at 20˚C (n = 251). Data collected by ALS; figure made by PMJ.

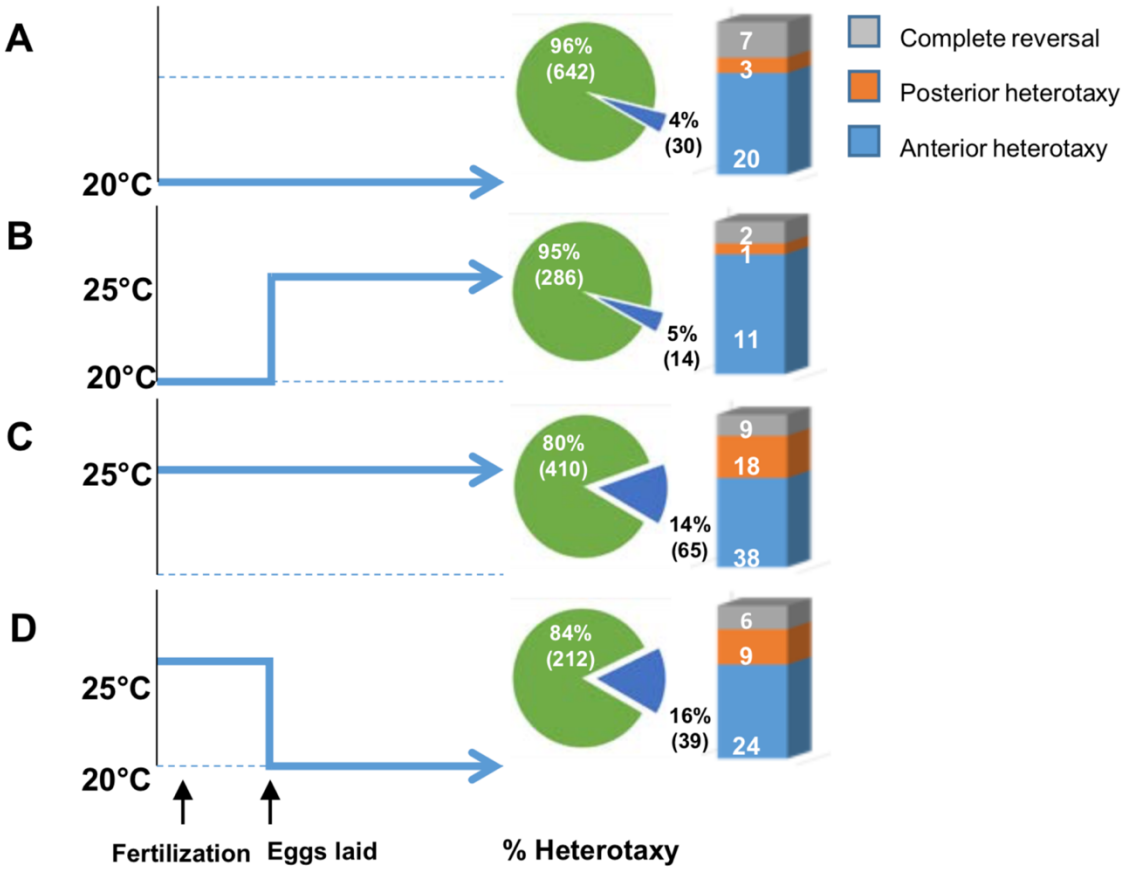

A
20°C
96%
(642)
4%
(30)
7
3
20
B
25°C
20°C
95%
(286)
5%
(14)
2
1
11
C
25°C
80%
(410)
14%
(65)
9
18
38
D
25°C
20°C
84%
(212)
16%
(39)
6
9
24
Fertilization
Eggs laid
% Heterotaxy
Complete reversal
Posterior heterotaxy
Anterior heterotaxy

Figure 4.6. Genome-Wide Association using Efficient Mixed-Model Analysis (EMMA) identifies 3 major genomic regions that are significantly associated with reversals in adult males across *C. elegans* isotypes.
$\log_{10}$ of *p-values* for each SNP were calculated using likelihood ratio tests and maximum-likelihood estimation. To control for multiple comparisons, a Bonferroni cut-off was used and is represented by the black line. Genome locations are represented on the x-axis, chromosomes indicated by the numbered boxes (I-V, X), and position along the chromosome is represented in mega-base pairs (mb) on the x-axis. Figure and analysis by MRA based on a script provided by Thomas Turner and modified for *C. elegans* by YTC.

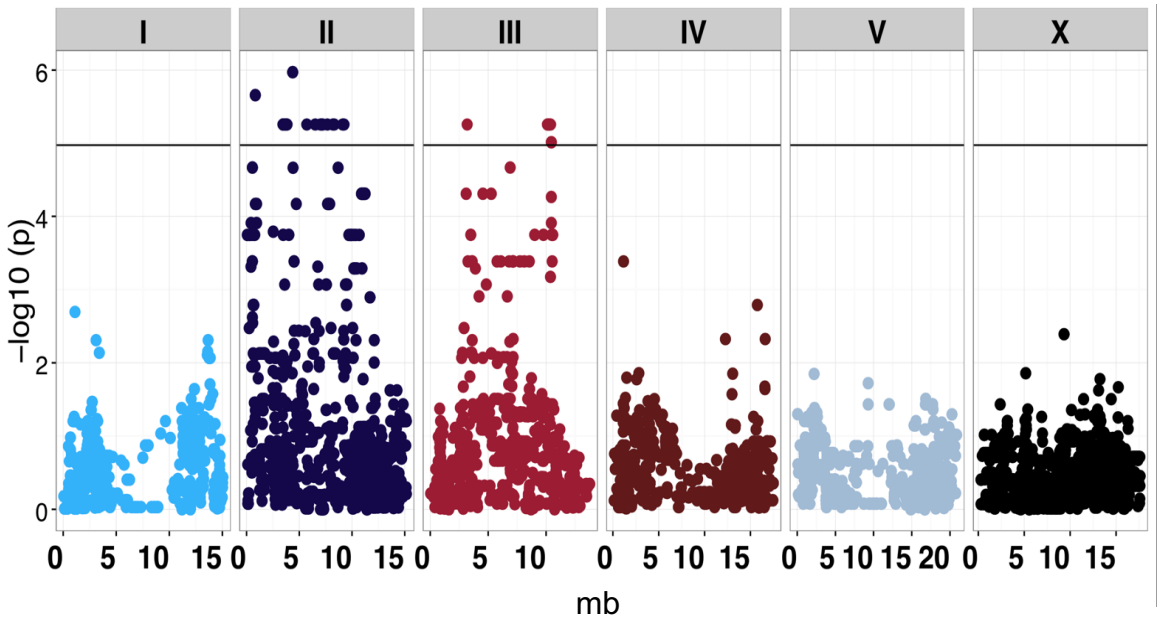

I
II
III
IV
V
X
−log10 (p)
0
2
4
6
0 5 10 15
0 5 10 15
0 5 10
0 5 10 15
0 5 10 15 20
0 5 10 15
mb

Figure 4.7. Frequency of reversals in males of recombinant inbred lines (RILs).
Each of 72 RILs is represented by a dot. The y-axis corresponds to % reversals. Parental strains are marked by dashed lines (AB4 in blue and QX1211 in red). Sample size varies from 30 to 100 males per RIL. Data collected by DCC and MRA.

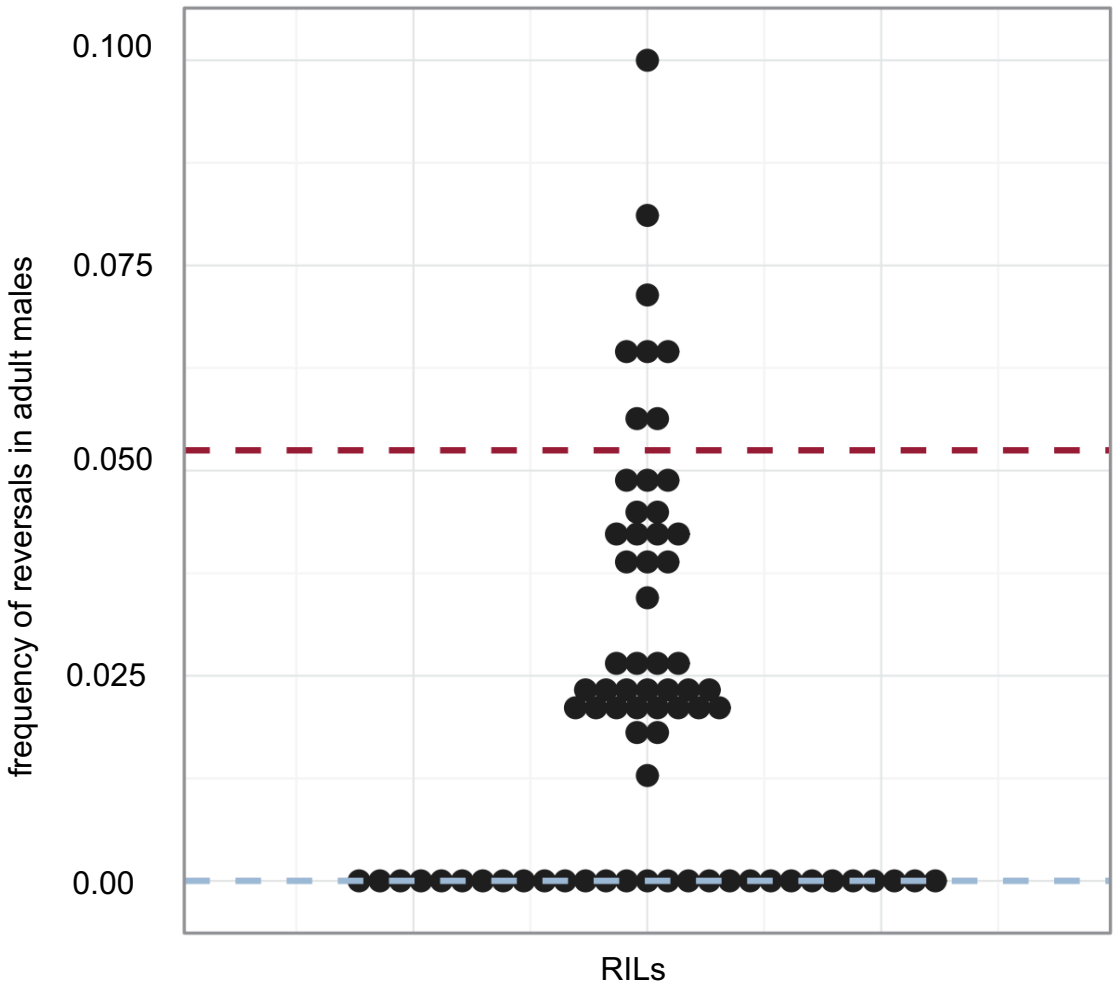
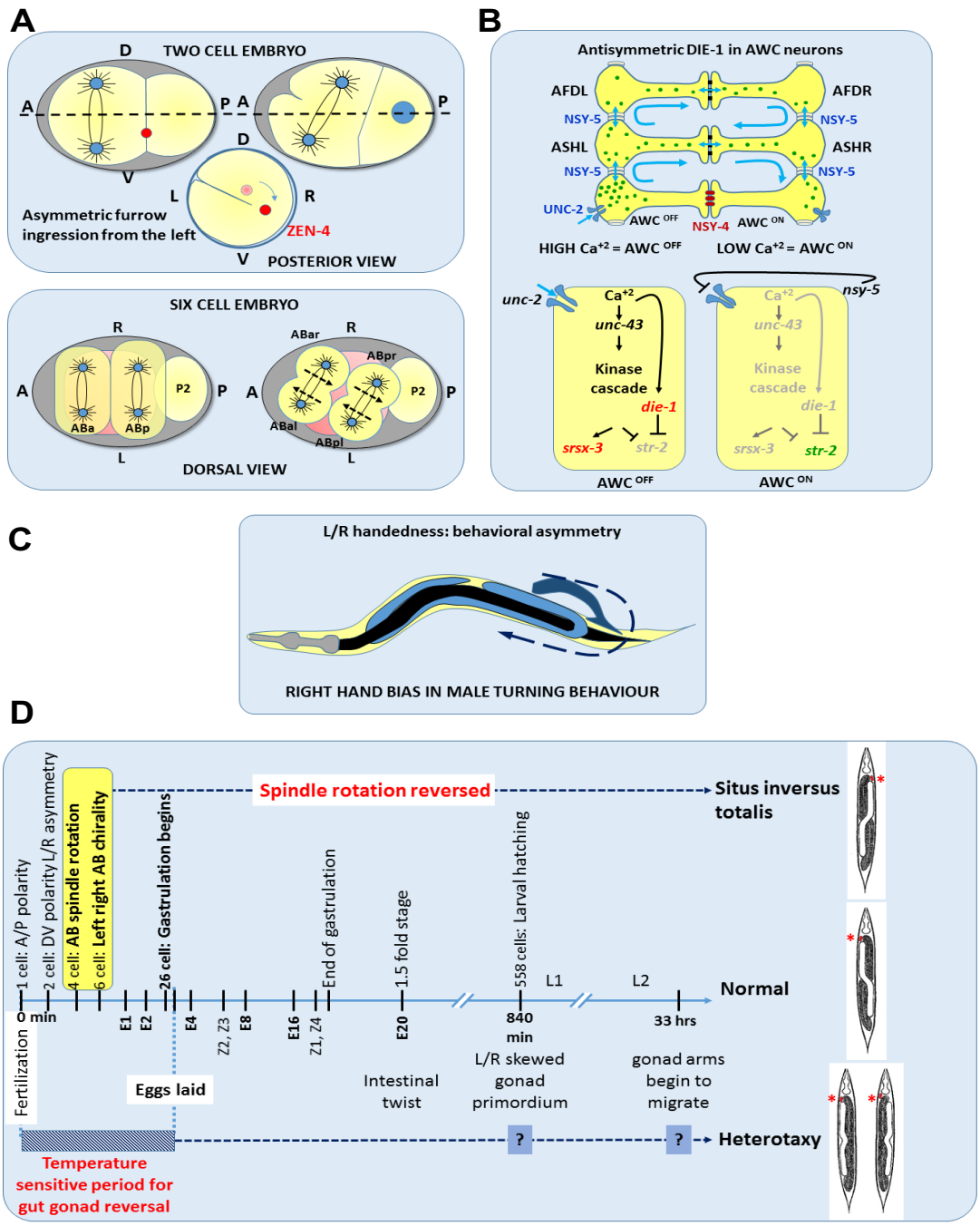

0.100
0.075
0.050
0.025
0.00
frequency of reversals in adult males
RILs

Figure 4.8. Left-right developmental and behavioral asymmetries in *C. elegans.*

(A) Actomyosin cortical flow-driven bilateral asymmetry in the early embryo. The one-cell embryo exhibits a dextral rotation within the eggshell as a result of chiral flows generated by cortical actomyosin contractility. The chirality is maintained in the AB cell resulting in left-right differences in the cortex, which is manifested in asymmetric cleavage furrow initiation from the left side (167). During the transition from 4- to 6-cell stage, the torque generated by counter-rotating cortical flows in the two AB daughter cells results in skewing of the spindles, resulting in a chiral arrangement of the four granddaughters of the AB blastomere . This chirality establishes most of the later L/R anatomical asymmetry of the animal (75). (B) Calcium signaling through gap junction coupled cell network biases asymmetric neuronal fate. The L/R differences between the two, functionally distinct, bilateral AWC neurons are stochastically determined. High calcium signaling in AWC autonomously promotes AWC^{OFF} (208, 211, 212) fate via the expression of the transcription factor DIE-1 (182), while calcium signaling in the non-AWC NSY-5 gap junction cell network promotes AWC^{ON} fate. In 50% of the animals AWC^{ON} is on the right and AWC^{OFF} on the left; in the other 50%, their positions are reversed. (C) Bilaterally asymmetric motor behavior. Males exhibit a right-hand turning bias that is independent of left-right anatomical asymmetry and embryonic chirality (118). (D) Gut/gonad heterotaxy: L/R embryonic chirality establishes asymmetric cellular interactions that result in bilateral differences in cell and organ positions. In normal (dextral) worms the anterior coelomocytes (marked by asterisks) and the anterior gonad arm of the hermaphrodites and entire gonad in males is on the right while the gut is on the left. Reversal of spindle skewing of the AB

daughter cells results in animals with completely reversed handedness (*situs inversus totalis*) (85). The gut/gonad reversals (anterior and posterior heterotaxy) observed in many natural isotypes shows temperature-sensitivity during early embryonic development, which affects L/R asymmetric migration of gonad arms later during larval development. They are independent of the initial L/R symmetry breaking events in the embryo. Figure made by PMJ.


A
TWO CELL EMBRYO
D
A
P
V
L
R
Asymmetric furrow ingression from the left
ZEN-4
POSTERIOR VIEW
SIX CELL EMBRYO
ABar
ABpr
ABa
ABp
ABal
ABpl
P2
DORSAL VIEW
B
Antisymmetric DIE-1 in AWC neurons
AFDL
AFDR
ASHL
ASHR
NSY-5
UNC-2
AWC OFF
NSY-4
AWC ON
HIGH Ca+2 = AWC OFF
LOW Ca+2 = AWC ON
unc-2
Ca+2
unc-43
Kinase cascade
die-1
srsx-3
str-2
nsy-5
C
L/R handedness: behavioral asymmetry
RIGHT HAND BIAS IN MALE TURNING BEHAVIOUR
D
1 cell: A/P polarity
2 cell: DV polarity L/R asymmetry
4 cell: AB spindle rotation
6 cell: Left right AB chirality
26 cell: Gastrulation begins
End of gastrulation
1.5 fold stage
558 cells: Larval hatching
L1
L2
Spindle rotation reversed
Situs inversus totalis
Normal
0 min
E1
E2
E4
Z2, Z3
E8
E16
Z1, Z4
E20
840 min
33 hrs
Fertilization
Eggs laid
Intestinal twist
L/R skewed gonad primordium
gonad arms begin to migrate
?
Heterotaxy
Temperature sensitive period for gut gonad reversal

Chapter Five

Future directions

Introduction

Since the first original collection of *C. elegans* isotypes were sequenced and publicly available to the *C. elegans* community (144) different groups have used quantitative genetics to identify causal genetic variants for phenotypic variation in a wide variety of biological processes including but not limited to specific mating behaviors, copulatory plugs, pathogen resistance, and vulva cell patterning (10, 41, 43, 45, 213, 214). Here we attempted to use quantitative genetics to identify causal variants in or gene regulatory networks involved in error-correction during development. To do so, we quantified as many traits as we could find that naturally varied in 85-100 *C. elegans* isolates (data collected by DCC, MRA, DT, PMJ, YTC and ALS, number of isolates vary with phenotype datasets) and found two isolates that were consistently highly variable or show no or significantly less phenotypic variance for all traits measured, such length of newly hatched L1 larvae, the number of stem-cells in the proximal region of the distal-tip cells of the hermaphrodite germline (unpublished, ask PJ for further data), requirement for SKN-1 in endoderm development, arrangement of major internal organs for both sexes (Chapter 4), and stochastic ray loss in the adult male tail (Chapters 2 and 3). We reasoned that because there seemed to be a consistent trend of high and low variability between these two highly sequentially divergent *C. elegans* isolates AB4 (a strain belonging to the CB4858 haplotype group) and QX1211 (40) there we may be able to identify a genetic mechanism (s) for error-correction during development, potentially similar to Hsp90 (7). We made recombinant inbred lines from these two strains (AB4 and QX1211) and used GWAS and QTL mapping to identify any and all QTL associated with any

of these phenotypes. Computational methods were able to identify loci associated with stochastic ray loss in the male tail (Chapter 3) and heterotaxy of the gut and gonad in males (Chapter 4). While unfortunately there was no GRN known to be affecting heterotaxy (at least in *C. elegans* to the best of our knowledge, Chapter 4), future work will require screening near isogenic lines with small regions of associated QTL to narrow associated regions and/or variant effect prediction analysis to identify candidate genes. In chapter 2, we showed that stochastic ray loss in N2 is primarily driven by the evolutionarily conserved pro-apoptotic factor EGL-1, however residual defects in caspase (CED-3 or CSP-1A) or caspase activator (CED-4) suggest there is potentially another caspase (s) or genetic factors involved in this type of stochastic cell death. To identify these factors, we used standing genetic and phenotypic variation in the propensity for ray defects in males from 87 unique *C. elegans* isotypes. While GWAS was able to identify one locus on chromosome II, recombinant inbred lines identified two more QTL on chromosomes III and X that explain ~55% of natural genetic variation (Chapter 3). While near isogenic lines were only able to exclude a small region on chromosome II, these and analysis of F1 males from reciprocal crosses reveal complex epistatic interactions between loci. As shown in chapter 4, we found that two QTL are associated with heterotaxy in males, but we were unable to identify variants for the phenotype in hermaphrodites, suggesting that potentially a sex-specific mechanism(s) may be involved, but it is also quite possible that any significant associations are confounded by population structure or there is not enough statistical power (despite the sample-size) to detect any significant SNPs in *C. elegans* isotypes or recombinant inbred

lines made between isotypes N2 and MY16. (data unpublished, see R scripts at https://github.com/RothmanLabCode).

Screening genes in QTL associated with stochastic apoptosis in the male tail

As presented in Chapter 2, while ~96% of ray defects were reduced in an *egl-1* null background, residual ray defects in caspase and caspase-activator null genetic backgrounds suggest that other caspases or regulators of caspase activity are involved. Interestingly, CSP-1A has shown to have a small reduction in ray defects compared to N2 and removes the right-handed bias in N2 and *ced-3(n717)lof* backgrounds, but another caspase-like protein CSP-3 also regulates aCED-3 *in vivo* (in the soma) and more alleles of *csp-1* that affect the active-site, etc. of CSP-1A should be screened in order to decipher the specific role of CSP-1A in stochastic apoptosis. Due to the difficulty in making near-isogenic lines for chromosomes II and X, most likely due to genetic incompatibilities (215) between two isolates AB4 and QX1211, only a small region on chromosome II was able to be eliminated as the small region of AB4 in an otherwise assumed to be QX1211 background as ray defects were not suppressed in this background when compared to QX1211 (Chapter 3). As an alternative, variant effect prediction software (VEP) was used to generate a list of candidate genes for some kind of screen with the original GBS data used for QTL mapping. In addition to identifying potentially high impact variants (i.e. non-synonymous amino acid substitutions, splice-variants, insertions/deletions (indels)) that would affect expression or potentially protein function, insertions/deletions or changes in restriction enzyme sites can also be used for genotyping NILs without sequencing as Sanger sequencing the PCR products of 3 or 4 different markers from

hundreds of males was the most expensive and time consuming part of making NILs from wild isolates (i.e. without fluorescent or physical markers available in N2). These candidate genes (all on chromosomes II, III and X) are listed in Table 3.1. Variants were filtered based on the number of predicted high-impact and then moderate impact mutations. These candidates were fed to WormBase ParaSite (216), where genes were further filtered based on known associated phenotypes – specifically accumulated somatic cell corpses. Interestingly, most of the variants were on chromosome III, so perhaps the next NIL to be made should be using this chromosome instead (if possible). Ideally, a NIL would confirm a narrow region and the CRISPR/Cas9 (217) system could be used to edit single genes in either parental background.

Additionally, the lab of Dr. Oliver Hobert (Columbia University, Department of Biological Sciences) has found variable numbers of male-specific interneurons from the B and U lineages – EF and DX GABA interneurons (49). In the original lineage analysis by Sulston (49) there are dashed lines at the terminal step for differentiation of these neurons, suggesting that the lineage for that particular step is unclear and to be determined. Oliver suggested that the variable numbers of interneurons seen via GABA-antibody staining in N2 males may be occurring because of stochastic apoptosis in the terminal lineage (218). To test this, fluorescent marker strains with a mutation for the main engulfment pathway such as *ced-1(-)* along with a fluorescent marker that specifically to these interneurons (218) would allow for accumulation of corpses which can be identified by DIC microscopy (both strains are in the lab

collection). This would be a potentially interesting side-project and an honor to add to the lineage.

Using machine learning and the natural variation in *C. elegans* isotypes to identify candidate genomic regions: a comparison to GWAS and QTL mapping

Summary

Genome-wide association studies (GWAS) and quantitative trait loci (QTL) mapping have been successful in identifying genetic variants associated with complex traits in both humans and model organisms. However, in complex diseases or traits where multiple loci contribute, those that show small effects individually may be missed by these methods due to low statistical power from testing thousands of single-nucleotide polymorphisms (SNPs). Here we present a case-study on how machine learning algorithms, specifically ElasticNet regression (219–221), may be used as a complementary approach to standard mapping techniques to identify not only loci with major contributions to the phenotype but also variants that may have small effects on their own, but when considered together drive phenotypic diversity. We applied ElasticNet regression to identify loci that control the plasticity in the gene-regulatory network controlling endoderm development. We have previously found significant variation in 94 unique isotypes for the necessity of the maternal transcription factor, SKN-1, for formation of endoderm using *skn-1* RNAi (118). While the laboratory reference strain N2 shows a partially penetrant phenotype

where ~30% of embryos form endoderm, we have found in the *C. elegans* isotypes the variation ranges from 0 to ~60%. GWAS using linear mixed-model analysis (EMMA) identified a single peak on chromosome IV, while QTL mapping using RILs made from N2 and isotype MY16 (~2% endoderm formation) revealed several loci on chromosome I, II, IV and X. ElasticNet regression identified overlapping regions for the QTL on I, II, IV, X in addition to a locus on chromosome V (Spearman's rho, $R^2 = 0.55$, $p = 6.5x10^{-5}$). Current work involves testing novel regions, such as those on chromosome V and its effect on endoderm development in different genetic backgrounds'. Modified/updated abstract that was submitted to International Worm Meeting 2017 (see CV).

Introduction

Genome-wide association studies (GWAS) have been very successful in identifying variants associated with a myriad of diseases (a few of the many examples: (23, 222–225)). The continuing advancements in sequencing technology and decreasing costs has made data collection easier than ever, where now genetic counseling is nearly routine (226). However, analyzing large datasets with hundreds of thousands of SNPs to find the most informative ones in a meaningful way is not trivial and GWAS does have some limitations: there may be low statistical power where the effect-size may be small (by nature), conflation with population structure (such as the infamous 'chop-stick' genes (227)) and variability or inconsistencies between SNP-calling methods (228). QTL mapping studies even in basic model organisms have identified multiple loci with opposing effects on the same phenotype and univariate methods may not be

enough to parse these effects – such as heterotaxy in hermaphrodites (unpublished data, see Chapter 4). While there are many algorithms that have been developed to account for population structure such as those that include identity-by-state or descent (IBS or IBD) kinship value (39), the first issue of identifying relevant features remains a challenge. Many studies have used different forms of machine learning, such as LASSO or Ridge Regression, where SNPs are treated as potentially predictive features and identified the same genetic loci identified by traditional methods like QTL mapping (229–234).

We chose ElasticNet algorithm from scikit-learn (235), as it has the best features of both Ridge Regression and LASSO algorithms. In Ridge Regression, L1 normalization is used shrink regression coefficients of the least informative features towards 0,but does not remove any features increasing the likelihood of overfitting. LASSO has a harsher penalty term which helps prevent overfitting, but equally informative features are chosen randomly to shrink the number of parameters and it's not clear how LASSO chooses the representative feature. ElasticNet uses both forms of regularization and in sckit-learn this is called the L1-L2 ratio. When this ratio is either 0 or 1, the algorithm performs like Ridge Regression or LASSO respectively (219, 220, 232, 236, 237). This flexibility allows for identifying the number of features and penalty-term (L1-L2 ratio) to be empirically determined. However, to the best of our knowledge novel regions of interest identified by machine learning have not been confirmed *in vivo*.

Here we show that *C. elegans* could be a system where results from machine learning methods can be directly screened in a living organism, specifically in a model that mimics the same levels of genetic variability that exists in human populations and has an extensive genetic toolkit and highly reproducible lineage and life cycle (40, 49, 50, 145, 238–240). In particular, the genetic toolkit of *C. elegans* now includes a large, genetically diverse collection of fully sequenced *C. elegans* isotypes (approximately 330 unique haplotypes as of this publication) that have been used to map a number of quantitative traits, in addition to a wealth of high throughput phenotyping methods (46). And while *C. elegans* may only have ~1000 cells, the lineage is known, and the molecular biology techniques are available to study mechanisms at the cellular level for any causal genes. To test how well ElasticNet regression identifies gene(s) of interest in *C. elegans*, we use the previously published significant variation in the dependency for SKN-1 in endoderm specification gene regulatory network in 94 unique *C. elegans* isotypes and their associated QTL. Previous work has illustrated that there is natural variation in the requirement for the maternal protein SKN-1, a bZip transcription factor and ortholog to mammalian Nrf transcription factors (118, 194). For specification of endoderm in the *C. elegans* E blastomere, high levels of either SKN-1 or POP-1, a Wnt ligand, are sufficient for specification of endoderm (155, 241). However, depending on the genetic background of the strain, endoderm specification will vary. Upon removal of *skn-1* by RNAi, between 0.9% and 60% of embryos will specify endoderm, depending on their isotype. Additionally, by using the endoderm specification as an example, results from ElasticNet can be verified by comparison to regions identified by standard association mapping techniques and *in vivo.*

We found that the regions identified by machine learning – specifically ElasticNet regression, all overlap with a region identified by one or more methods like EMMA and QTL mapping of RILs, suggesting that machine learning can identify cryptic genetic variants as well, in addition to novel regions on chromosomes V and X (in addition to regions already identified). Further work involves confirming the region identified on chromosome V by introgression of a region from the laboratory reference strain into strains with high and low dependency on SKN-1, in addition improving the modeling system with better ways for feature selection or a way to include a kinship matrix of some kind (38). It is possible that the novel regions on V and X are due to spurious associations from population structure, as GWAS (with no kinship matrix) identified a region on chromosome V but was not considered significant using EMMA (with kinship matrix). It is also necessary to determine, using published QTL data verified *in vivo* or previous work in the reference strain N2 (i.e. previously published information), the false positive and negative rates. However, for traits where EMMA or QTL mapping did not find any significant associations, it may be possible to try ElasticNet regression as a way to cast a wide net, and by using a model organism that is genetically tractable like *C. elegans,* candidates can be screened relatively efficiently (depending on the phenotype).

Results

Schematic for ElasticNet Regression modeling

Machine learning is a powerful tool for multi-variate analysis when the number of features (predictors) is greater than the number of samples in the dataset. This is

especially the case in genomics. Here our goal is to discover a function that maps SNPs to trait frequency – the fraction of endoderm being made in embryos of a specific isotype when mothers are fed *skn-1* RNAi using n = 4690 binary SNPs and 100 *C. elegans* natural isolates. To identify loci, we followed the schematic seen in Figure 5.1. Before training, we chose to remove SNPs with low variance for two reasons: we are interested in highly variable SNPs (as potential sources for phenotypic variation) and to reduce the initial number of features before feature-selection. This is arbitrary though and is entirely based on the goal of the study and empirical data. We also needed to consider the form of cross-validation, feature-selection (i.e. the number of features for modeling or how to select them), in addition to the penalty-term for optimization to avoid over-fitting (236, 242). Cross-validation is one of the major advantages of machine learning in that significant features are used to predict the outcome or phenotype of a left out or unseen dataset (test-set, Figure 5.1). This can be repeated using different subsets of data to tune regression coefficients (232). We chose leave-one-out cross-validation method to avoid having to ensure similar distributions of training-sets during cross-validation. Additionally, we optimized the number of features and L1-L2 ratio used by ElasticNet. The L1-L2 ratio is the penalty term is used to reduce all regression coefficients near zero, leaving the most informative features with the highest, positive regression coefficients (230).

ElasticNet Regression identifies regions identified by QTL mapping and GWAS

Previous work has identified regions on chromosomes I, II, IV and X (155). We used the genetic and phenotypic variation (and same SNP map used for GWAS) in 94 *C.*

elegans isotypes and ElasticNet Regression (Figure 5.1) to identify loci associated with SKN-1 dependency and defined possible regions of interest using linkage-disequilibrium. We chose optimal number of features and L1-L2 penalty term based on 94 rounds of leave-one-out cross-validation for every combination of L1-L2 ratio between 0 and 1 ([0,1,0.1] and n = 250,500 750,1000, 2000 SNPs per training set). Spearman's rho was calculated using results from the 94 test-sets (per L1-L2 ratio value and N number of features) to demonstrate how well ElasticNet was able to correlate phenotype with genotype on testing sets after training. ElasticNet regression built a model using n = 1431 SNPs, L1-L2 ration = 0.3 that are significantly associated with the phenotype (Spearman's rho, R^2= 0.55, p = $6.5x10^{-5}$). In order to test for significance of SNPs chosen during optimization, we used n = 1000 permutations to build 95% confidence intervals with optimal L1-L2 value and number of features (Figure 5.2). SNPs were considered significant if their lower-bound (mean – 2 X standard-deviations) was greater than the upper-bound of the 95% confidence interval calculated during permutations. Based on these criteria, we identified n = 15 significant SNPs (Figure 5.2). We then used PLINK to calculate linkage disequilibrium of the significant SNPs to define possible regions and these were then compared to the regions identified by GWAS with using the EMMA model and 95% confidence intervals from QTL mapping (Figure 5.3). We found that ElasticNet identified a novel region on chromosome V, in addition to two smaller QTL on X.

Effect plots of SNPs on chromosome V suggest a novel region for confirmation testing *in vivo*

While ElasticNet identified n = 1431 SNPs as significant, only 15 SNPs considered significant by permutation testing and 95% confidence intervals. We then compared the SNPs considered significant by all models (Figure 5.3) and found that ElasticNet regression identified SNPs from both GWAS using *C. elegans* isotypes and QTL mapping made with recombinant inbred lines from two isotypes MY16 and N2. While all methods identified a region on chromosome IV, ElasticNet regression identified a novel region on chromosome V. While this region was identified using GWA without a kinship matrix, it was no longer significant when a kinship matrix (IBS) was included. It could very well be this association may be due to spurious associations from population structure, but it is still possible that this QTL is biologically relevant. We recorded the number of times each SNP was included in the modeling process – as we thought it would be a way to discern between SNPs that were significant during only one round of cross-validation, I .e. for one isotype – as a way to avoid spurious associations. Either way – it is informative about how to apply machine learning methods for this type of problem. It is possible to test this using near-isogenic lines using naturally varying *C. elegans* isotypes and to identify candidate strains, we plotted the phenotype frequency vs. interaction of alleles at significantly associated SNPs (n = 3 SNPs, ANOVA, *dof* = 93, $p = 3.23 \times 10^{-5}$) and found that there was a significant difference between allele pattern CCC and TTA (Tukey's adjusted p-values, p = 0.0416) (Figure 5.4). This suggests that potentially

isotypes with CCC and TTA allele combinations at the three significant SNPs identified by ElasticNet may be candidates for near isogenic lines.

Discussion

ElasticNet and its closely-related algorithms such as Ridge Regression and Lasso have been successful in classifying cancers, improve quality or select for certain traits in livestock and plants, annotate genomes or identify regulatory elements throughout a given genome, and to predict features of compounds that may prove useful in treating certain diseases (234, 242–245). To the best of our knowledge, genomic regions considered significant by machine learning methods have not been tested in a living organism. This presents a significant hurdle for using machine learning in patient diagnostics or personalized medicine, as these methods may find statistically significant SNPs or features, but it is not clear how changing this or any region would affect overall development or viability. Here we propose that *C. elegans* is an attractive model for testing the effect of SNPs not only on their associated phenotype but also on development as a whole in different genetic backgrounds. However – the current method as it is presented is not complete and significant tasks remain. For example, we chose to remove SNPs with low genetic variance before training as we are interested SNPs associated with genetic variability and it may be a way to simplify an already complicated model and improve accuracy. However – this step is arbitrary, and the type of feature selection may not be applicable to all datasets. Additionally, the current model does not account for population structure and any results considered statistically significant require further thought from experimentalists. It is

also not clear how or when to include linkage-disequilibrium into the modeling process. We chose to identify single significant SNPs using 95% confidence intervals and calculated linkage disequilibrium of these SNPs to identify potential regions of interest (Figure 5.4). ElasticNet was able to identify regions identified using recombinant inbred lines and GWAS in addition to novel ones but the true test would be to test a novel region, such as one on X or V (Figures 5.3, 5.4), using different isotypes.

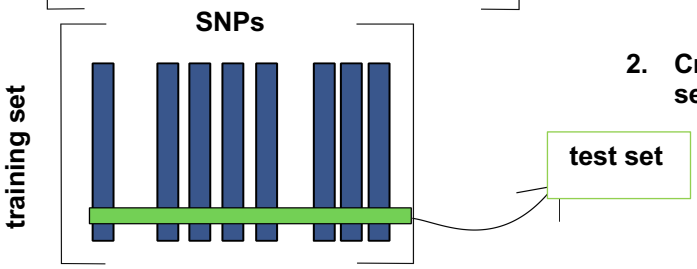
Materials and methods

To optimize conditions for the original data-set, we performed 95 rounds of leave-one-out cross-validation using all combinations of L1 penalty values (ranging from 0 to 1) and significant number of features to be kept after feature selection (n = 250, 500, 750, 1000, 2000), leading to a total of 1960 iterations. To choose SNPs for the model, we used the most significant features associated with the phenotype by Wilcoxon-Rank Sum tests (without correcting for multiple comparisons). During the optimization process, the N-most significant SNPs are fed to ElasticNet and regression coefficients (β) are stored in addition to how many times each SNP was included in model selection. Optimal values were chosen during the training/testing of the original data set using Spearman's *rho* to test how well the model predicted the unknown phenotype. For this dataset, optimal L1 value is 0.9 and the optimal number of initial features fed to ElasticNet = 1000. These parameters were then used in permutation-testing in order to build 95% confidence intervals (n = 1000). For both optimization and permutation testing, we calculated 95% confidence intervals for regression coefficients of each SNP. To consider which SNPs were

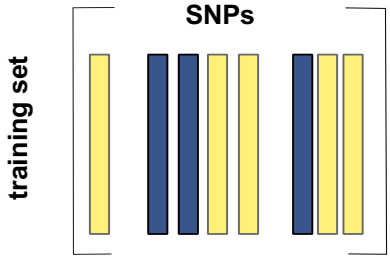
significant, we chose SNPs that were considered significant during optimization in more than half of the rounds of cross-validation (n = 98/2 = 49) and whose confidence intervals did not overlap with those from permutations. After we identified significant SNPs, we used linkage disequilibrium calculated in PLINK v1.9 to identify regions of interest.

Figure 5.1. Scheme for ElasticNet regression.
We decided to use ElasticNet Regression, as it has both attractive features of Lasso and Ridge Regression. Data analysis follows the following pipeline: 1) We removed those with variance < 0.2, reasoning that these SNPs are most likely not contributing to phenotypic variation, but this step is arbitrary. 2) Machine learning can be used to predict a value, and then update its model based on the strength of the prediction. In order to do so, depending on the style of cross-validation (either k-means or leave-one-out), a testing set must be removed from the remaining data. 3) Feature selection – we chose SNPs that were most significantly associated with the phenotype using student's t-tests with adjustment for un-equal variance and without adjustment for multiple comparisons, but this can be done using other statistical tests. 4) Training: SNPs most strongly associated with the phenotype and isolates remaining in the dataset is fed to ElasticNet regression model, where a regression coefficient is calculated for each SNP. Testing: using regression coefficients and significant SNPs from training data, ElasticNet regression predicts the phenotype of the strain left out before feature selection (testing set). Record coefficients for each SNP and repeat process with new training and testing sets.

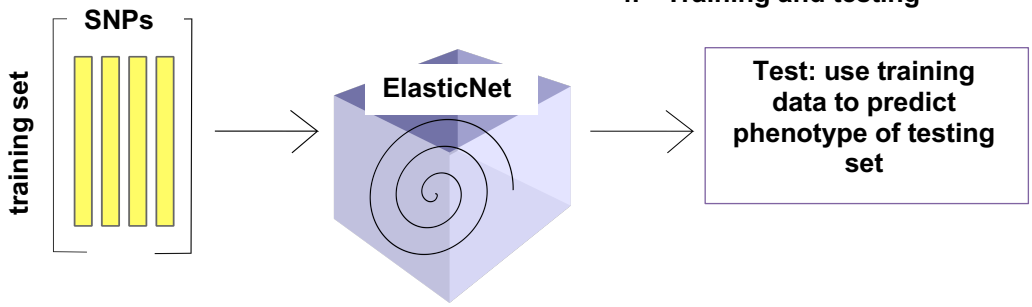
1. Pre-process data. We removed SNPs with low variance.



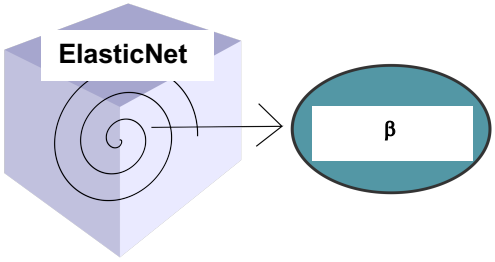
2. Create training and testing sets for cross-validation.



3. Feature selection: choose n-best features for Elasticnet regression modeling based on association with phenotype.



4. Training and testing



5. Store regularized coefficients (beta).

Figure 5.2. Regularized regression coefficients.
Manhattan style-plot with position of SNP in mb on the X-axis and average coefficient from regression (regularized) on the y-axis. Dots represent the average coefficient and black bars represent 95% confidence intervals calculated using Wilcoxon Rank Sum tests using data from n = 1000 permutations. In order to identify significant regions, the initial dataset is analyzed using the scheme above. To separate signal from noise, the phenotype is permuted 1000X and using the same number of features, penalty terms (L1 and L2) determined by 100-rounds of leave-one-out (LOO) cross-validation of original dataset.

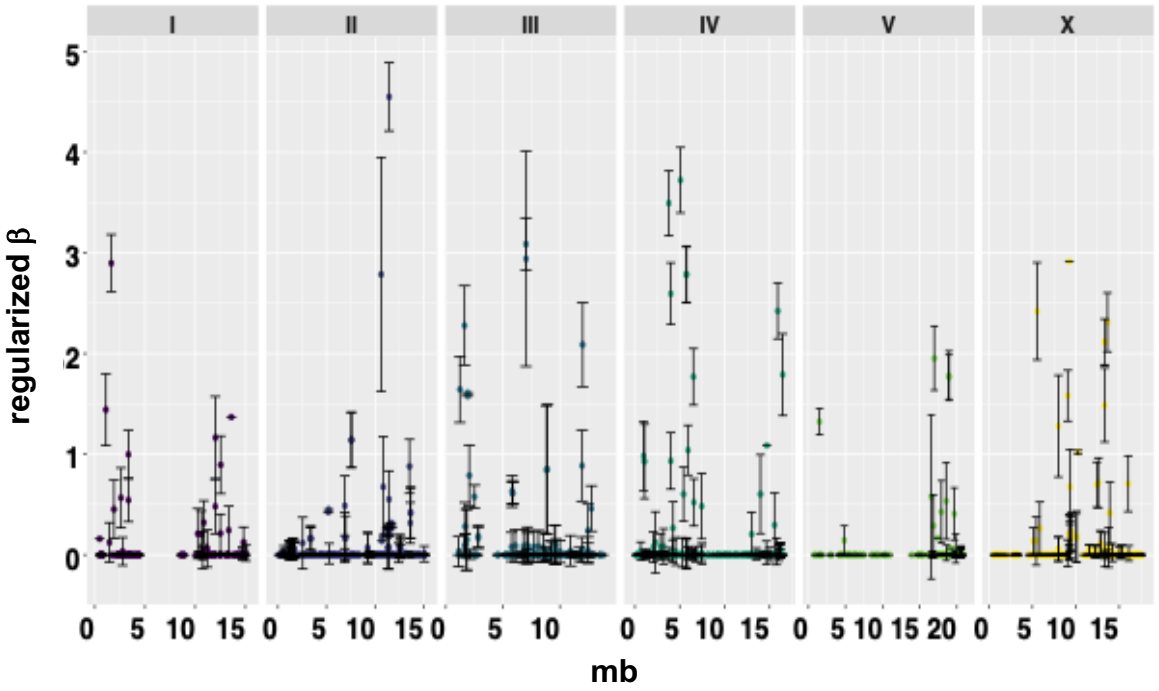

I
II
III
IV
V
X
regularized β
mb

Figure 5.3. Comparison of regions identified by several models.
An overlay of the physical regions identified using two different sources of genotypic variation: GWAS analysis using genotypic variation of wild isolates and QTL mapping using genetic variation associated with RILs made from the reference strain N2 and isolate MY16 (155). Black bars represent 2 mb sections of the chromosome.

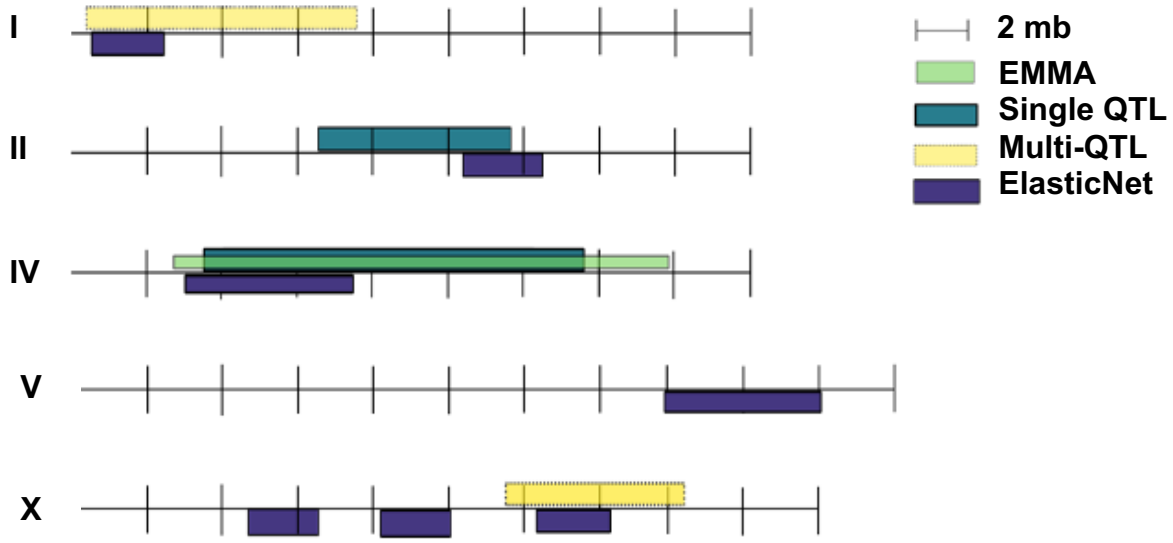
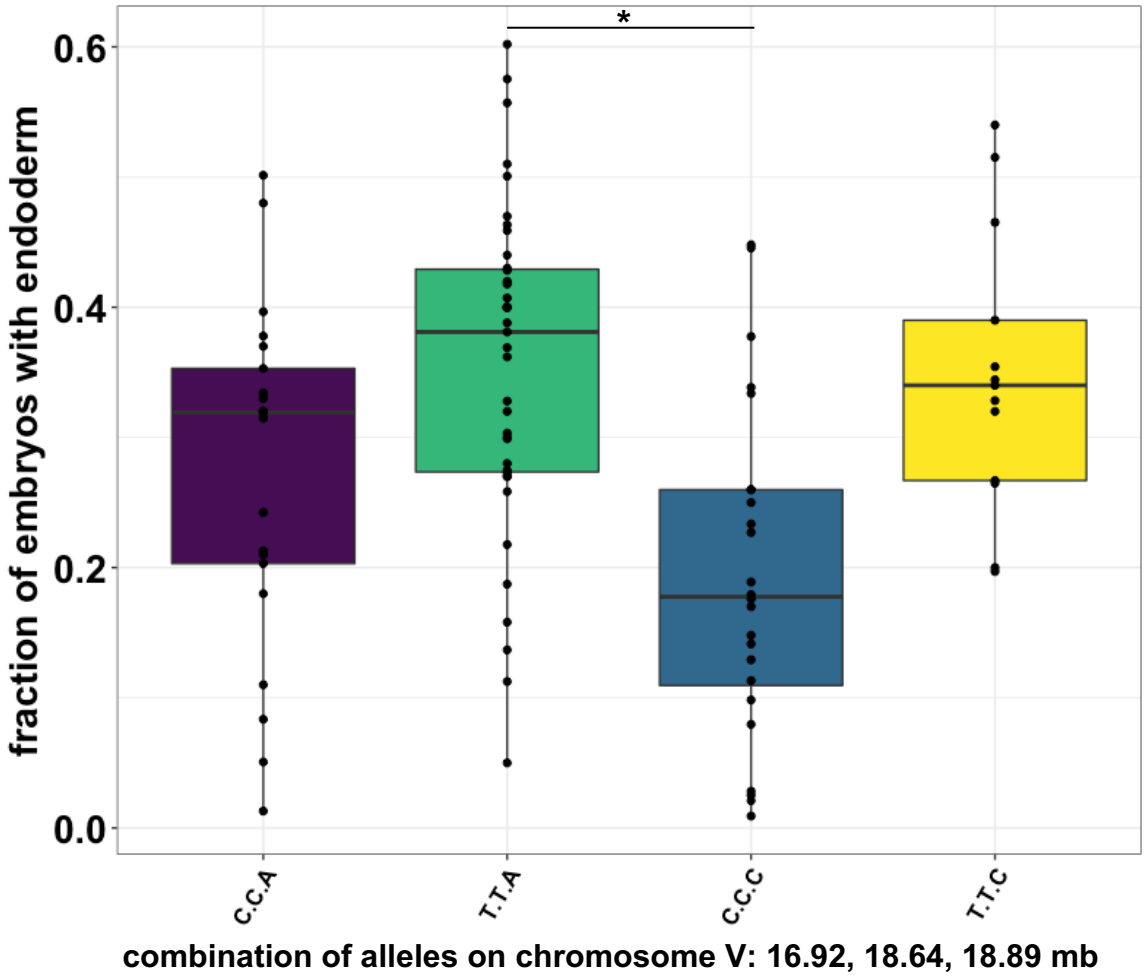

I
II
IV
V
X
2 mb
EMMA
Single QTL
Multi-QTL
ElasticNet

Figure 5.4. Effect plot of significant SNPs identified by machine learning on chromosome V.
Significant regions determined by permutation testing on chromosome V were found using ElasticNet Regression and GWAS – without accounting for population structure (as opposed to EMMA, which does). Three combinations of genotypes were compared against each other using linear model and multiple comparisons were corrected for using Tukey's method. * $p < 0.05$, no bar or marking indicates comparison was not significant ($p > 0.05$). Based on the allele patterns, a potentially interesting near-isogenic line would be made with an isolate with the genotype allele pattern of TTA and CCC at the SNPs identified using ElasticNet Regression on chromosome V. Two such isolates are JU1491 (fraction with endoderm = 0.000916, allele pattern = CCC) and JU440 fraction with endoderm = 0.602, allele pattern = TTA). By crossing these two strains with N2 (fraction with endoderm = 0.302, allele pattern ACC), the cross with JU1491 may reveal effect of the SNP or region around the SNP at ~16.97 Mb and the cross with JU440 may isolate and reveal the effect of the SNPs near 18.946 Mb.


*
0.6
0.4
0.2
0.0
fraction of embryos with endoderm
C.C.A
T.T.A
C.C.C
T.T.C
combination of alleles on chromosome V: 16.92, 18.64, 18.89 mb

References

1. C. H. Waddington, Genetic assimilation of an acquired character. *Evolution* **7**, 118–126 (1953).

2. C. H. Waddington, Canalization of Development and the Inheritance of Acquired Characters. *Nature*, 563–5 (1942).

3. R. A. Ganai, E. Johansson, DNA Replication-A Matter of Fidelity. *Mol. Cell* **62**, 745–755 (2016).

4. S. E. Tsutakawa, J. Lafrance-Vanasse, J. A. Tainer, The cutting edges in DNA repair, licensing, and fidelity: DNA and RNA repair nucleases sculpt DNA to measure twice, cut once. *DNA Repair (Amst)* **19**, 95–107 (2014).

5. C. S. Francklyn, DNA Polymerases and Aminoacyl-tRNA Synthetases: Shared Mechanisms for Ensuring the Fidelity of Gene Expression. *Biochemistry* **47**, 11695–11703 (2008).

6. S. L. Rutherford, S. Lindquist, Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).

7. C. Queitsch, T. A. Sangster, S. Lindquist, Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624 (2002).

8. N. Rohner, *et al.*, Cryptic Variation in Morphological Evolution: HSP90 as a Capacitor for Loss of Eyes in Cavefish. *Science* **342**, 1372–1375 (2013).

9. H. Shin, D. J. Reiner, The Signaling Network Controlling C. elegans Vulval Cell Fate Patterning. *J Dev Biol* **6** (2018).

10. M.-A. Félix, Cryptic Quantitative Evolution of the Vulva Intercellular Signaling Network in Caenorhabditis. *Current Biology* **17**, 103–114 (2007).

11. M. Barkoulas, J. S. van Zon, J. Milloz, A. van Oudenaarden, M.-A. Félix, Robustness and Epistasis in the C. elegans Vulval Signaling Network Revealed by Pathway Dosage Modulation. *Developmental Cell* **24**, 64–75 (2013).

12. C. Braendle, M. A. Félix, Plasticity and Errors of a Robust Developmental System in Different Environments. *Developmental Cell* **15**, 714–724 (2008).

13. M.-A. Félix, M. Barkoulas, Pervasive robustness in biological systems. *Nat Rev Genet* **16**, 483–496 (2015).

14. A. Wagner, *Robustness and Evolvability in Living Systems* (Princeton University Press, 2013).

15. M. J. Sutherland, S. M. Ware, Disorders of left-right asymmetry: Heterotaxy and situs inversus. *American Journal of Medical Genetics, Part C: Seminars in Medical Genetics* **151**, 307–317 (2009).

16. T. Yokoyama, *et al.*, Reversal of left-right asymmetry: a situs inversus mutation. *Science* **260**, 679–682 (1993).

17. M. Martinez-Vicente, Autophagy in neurodegenerative diseases: from pathogenic dysfunction to therapeutic modulation. *Seminars in Cell & Developmental Biology* (2015) https:/doi.org/10.1016/j.semcdb.2015.03.005.

18. R. S. Kahn, *et al.*, Schizophrenia. *Nature Reviews Disease Primers* **1** (2015).

19. C. M. Sellgren, *et al.*, Increased synapse elimination by microglia in schizophrenia patient-derived models of synaptic pruning. *Nature Neuroscience* **22**, 374–385 (2019).

20. T. M. Karrer, *et al.*, Brain-based ranking of cognitive domains to predict schizophrenia. *Human Brain Mapping* **40**, 4487–4507 (2019).

21. D. Dong, H. R. Zielke, D. Yeh, P. Yang, Cellular stress and apoptosis contribute to the pathogenesis of autism spectrum disorder. *Autism Research* **11**, 1076–1090 (2018).

22. B. Carpiniello, *et al.*, Treatment-Resistant Schizophrenia: Genetic and Neuroimaging Correlates (2019) https:/doi.org/10.3389/fphar.2019.00402.

23. G. I. D. Zubicaray, N. K. Hansell, G. W. Montgomery, N. G. Martin, Discovery of Genes That Affect Human Brain Connectivity : a Genome-Wide Analysis of the. 1–11 (2013).

24. E. H. Davidson, D. H. Erwin, Gene regulatory networks and the evolution of animal body plans. *Science (New York, N.Y.)* **311**, 796–800 (2006).

25. E. H. Davidson, M. S. Levine, Properties of developmental gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 20063–6 (2008).

26. E. H. Davidson, *et al.*, A Provisional Regulatory Gene Network for Specification of Endomesoderm in the Sea Urchin Embryo. *Developmental Biology* **246**, 162–190 (2002).

27. I. S. Peter, E. H. Davidson, Assessing regulatory information in developmental gene regulatory networks. **114** (2015).

28. E. M. Erkenbrack, E. H. Davidson, I. S. Peter, Conserved regulatory state expression controlled by divergent developmental gene regulatory networks in echinoids. *Development* (2018) https:/doi.org/10.1242/dev.167288.

29. K. T.-H. Lin, G. Broitman-Maduro, W. W. K. Hung, S. Cervantes, M. F. Maduro, Knockdown of SKN-1 and the Wnt effector TCF/POP-1 reveals differences in endomesoderm specification in C. briggsae as compared with C. elegans. *Developmental Biology* **325**, 296–306 (2009).

30. S. Bhattacharya, Q. Zhang, M. E. Andersen, A deterministic map of Waddington's epigenetic landscape for cell fate specification. *BMC Syst Biol* **5**, 85 (2011).

31. M. V. Rockman, Reverse engineering the genotype-phenotype map with natural genetic variation. *Nature* **456**, 738–44 (2008).

32. X. Zhang, S. D. Bailey, M. Lupien, Laying a solid foundation for Manhattan - "setting the functional basis for the post-GWAS era." *Trends in Genetics* **30**, 140–149 (2014).

33. J. Davey, *et al.*, Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**, 499–510 (2011).

34. Z. Zhang, *et al.*, Mixed linear model approach adapted for genome-wide association studies. *Nature genetics* **42**, 355–360 (2010).

35. B. M. Bolker, *et al.*, Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution* **24**, 127–135 (2009).

36. A. L. Oberg, D. W. Mahoney, Linear Mixed Effects Models. *Methods in molecular biology :Topics in Biostatistics* **404**, 213–234 (2007).

37. , Kinship Coefficients and Covariance Between Relatives. *Biostatistics*.

38. B. Wang, S. Sverdlov, E. Thompson, Efficient Estimation of Realized Kinship from Single Nucleotide Polymorphism Genotypes. *Genetics* **205**, 1063–1078 (2017).

39. , Accuracy of genomic selection for a sib-evaluated trait using identity-by-state and identity-by-descent relationships (January 25, 2020).

40. D. E. Cook, S. Zdraljevic, J. P. Roberts, E. C. Andersen, CeNDR, the Caenorhabditis elegans natural diversity resource. *Nucleic Acids Research* **45**, D650–D657 (2017).

41. R. Ghosh, E. C. Andersen, J. a Shapiro, J. P. Gerke, L. Kruglyak, Natural Variation in a Chloride Channel Subunit Confers Avermectin Resistance in C. elegans. *Science* **335**, 574–578 (2012).

42. E. C. Andersen, J. S. Bloom, J. P. Gerke, L. Kruglyak, A Variant in the Neuropeptide Receptor npr-1 is a Major Determinant of Caenorhabditis elegans Growth and Physiology. *PLoS Genetics* **10** (2014).

43. M. F. Palopoli, *et al.*, Molecular basis of the copulatory plug polymorphism in Caenorhabditis elegans. *Nature* **454**, 1019–22 (2008).

44. A. Bendesky, *et al.*, Long-range regulatory polymorphisms affecting a GABA receptor constitute a quantitative trait locus (QTL) for social behavior in Caenorhabditis elegans. *PLoS genetics* **8**, e1003157–e1003157 (2012).

45. D. E. Cook, *et al.*, *The genetic basis of natural variation in Caenorhabditis elegans telomere length* (2016).

46. E. C. Andersen, *et al.*, A Powerful New Quantitative Genetics Platform, Combining Caenorhabditis elegans High-Throughput Fitness Assays with a Large Collection of Recombinant Strains. *G3: Genes|Genomes|Genetics* **5**, 911–920 (2015).

47. S. J. Cook, *et al.*, Whole-animal connectomes of both Caenorhabditis elegans sexes. *Nature* **571**, 63–71 (2019).

48. J. G. White, E. Southgate, J. N. Thomson, S. Brenner, The Structure of the Nervous System of the Nematode Caenorhabditis elegans. *Philosophical Transactions of the Royal Society B: Biological Sciences* **314**, 1–340 (1986).

49. J. E. Sulston, H. R. Horvitz, Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. *Developmental Biology* **56**, 110–156 (1977).

50. J. E. Sulston, E. Schierenberg, J. G. White, J. N. Thomson, The embryonic cell lineage of the nematode Caenorhabditis elegans. *Developmental Biology* **100**, 64–119 (1983).

51. R. Lints, D. H. Hall, WormAtlas Male Handbook - Neuronal Support Cells - Rays. *WormAtlas* (2005) https:/doi.org/10.3908/wormatlas.2.10 (January 27, 2020).

52. R. Lints, D. H. Hall, Male neuronal support cells, overview. *Wormatlas* (2009) https:/doi.org/10.3908/wormatlas.2.9.

53. K. S. Liu, P. W. Sternberg, Sensory regulation of male mating behavior in Caenorhabditis elegans. *Neuron* **14**, 79–89 (1995).

54. S. W. Emmons, Male development. *WormBook : the online review of C. elegans biology*, 1–22 (2005).

55. J. E. Sulston, D. G. Albertson, J. N. Thomson, The Caenorhabditis elegans male: Postembryonic development of nongonadal structures. *Developmental Biology* **78**, 542–576 (1980).

56. K. M. Boatright, G. S. Salvesen, Mechanisms of caspase activation. *Current Opinion in Cell Biology* **15**, 725–731 (2003).

57. B. T. Hyman, J. Yuan, Apoptotic and non-apoptotic roles of caspases in neuronal physiology and pathophysiology. *Nature reviews. Neuroscience* **13**, 395–406 (2012).

58. C. L. Cusack, V. Swahari, W. Hampton Henley, J. Michael Ramsey, M. Deshmukh, Distinct pathways mediate axon degeneration during apoptosis and axon-specific pruning. *Nature Communications* (2013) https:/doi.org/10.1038/ncomms2910.

59. B. Conradt, Y.-C. Wu, D. Xue, Programmed Cell Death During Caenorhabditis elegans Development. *Genetics* **203**, 1533–1562 (2016).

60. G. Wang, *et al.*, CED-4 CARD domain residues can modulate non-apoptotic neuronal regeneration functions independently from apoptosis. *Scientific Reports* **9** (2019).

61. B. Conradt, H. R. Horvitz, The C. elegans Protein EGL-1 is required for programmed cell death and interacts with the Bcl-2-like protein CED-9. *Cell* **93**, 519–529 (1998).

62. R. Nehme, *et al.*, Transcriptional upregulation of both egl-1 BH3-only and ced-3 caspase is required for the death of the male-specific CEM neurons. *Cell Death & Differentiation* **17**, 1266–1276 (2010).

63. Y.-Z. Chen, J. Mapes, E.-S. Lee, R. Robert Skeen-Gaar, D. Xue, Caspase-mediated activation of Caenorhabditis elegans CED-8 promotes apoptosis and phosphatidylserine externalization. *Nature Communications* **4**, 2726–2726 (2013).

64. E. Kuranaga, Beyond apoptosis: caspase regulatory mechanisms and functions in vivo. *Genes to Cells* **17**, 83–97 (2012).

65. J. Suzuki, D. P. Denning, E. Imanishi, H. R. Horvitz, S. Nagata, Xk-Related Protein 8 and CED-8 Promote Phosphatidylserine Exposure in Apoptotic Cells. *Science* **341**, 403–406 (2013).

66. D. Xue, H. R. Horvitz, Caenorhabditis elegans CED-9 protein is a bifunctional cell-death inhibitor. *Nature* **390**, 305–308 (1997).

67. A. Nakagawa, K. D. Sullivan, D. Xue, Caspase-activated phosphoinositide binding by CNT-1 promotes apoptosis by inhibiting the AKT pathway. *Nature Structural & Molecular Biology* **21**, 1082–1090 (2014).

68. K. D. Sullivan, A. Nakagawa, D. Xue, J. M. Espinosa, Human ACAP2 is a homolog of C. elegans CNT-1 that promotes apoptosis in cancer cells. **4101** (2016).

69. M. Thellmann, J. Hatzold, B. Conradt, The Snail-like CES-1 protein of C. elegans can block the expression of the BH3-only cell-death activator gene egl-1 by antagonizing the function of bHLH proteins. *Development (Cambridge, England)* **130**, 4057–4071 (2003).

70. J. Hatzold, B. Conradt, Control of Apoptosis by Asymmetric Cell Division https:/doi.org/10.1371/journal.pbio.0060084.

71. N. Mishra, H. Wei, B. Conradt, Caenorhabditis elegans ced-3 Caspase Is Required for Asymmetric Divisions That Generate Cells Programmed To Die. *Genetics* **210**, 983–998 (2018).

72. D. P. Denning, V. Hatch, H. R. Horvitz, Both the Caspase CSP-1 and a Caspase-Independent Pathway Promote Programmed Cell Death in Parallel to the Canonical Pathway for Apoptosis in Caenorhabditis elegans. *PLoS Genetics* **9**, e1003341–e1003341 (2013).

73. X. Geng, *et al.*, Inhibition of CED-3 zymogen activation and apoptosis in Caenorhabditis elegans by caspase homolog CSP-3. *Nature Structural & Molecular Biology* **15**, 1094–1101 (2008).

74. X. Geng, *et al.*, Caenorhabditis elegans caspase homolog CSP-2 inhibits CED-3 autoactivation and apoptosis in germ cells. *Cell Death & Differentiation* **16**, 1385–1394 (2009).

75. W. B. Wood, Evidence from reversal of handedness in C. elegans embryos for early cell interactions determining cell fates. *Nature* **349**, 536–538 (1991).

76. S. R. Naganathan, S. Fürthauer, M. Nishikawa, F. Jülicher, S. W. Grill, Active torque generation by the actomyosin cell cortex drives left–right symmetry breaking. *eLife* **3** (2014).

77. B. De, O. Oceanologique, A Conserved Role for the Nodal Signaling Pathway in the Establishment of Dorso-Ventral and Left – Right Axes in Deuterostomes. **53**, 41–53 (2008).

78. C. F. Chuang, M. K. VanHoven, R. D. Fetter, V. K. Verselis, C. I. Bargmann, An Innexin-Dependent Cell Network Establishes Left-Right Neuronal Asymmetry in C. elegans. *Cell* **129**, 787–799 (2007).

79. A. Asan, *et al.*, Morphogenesis of the C. elegans Intestine Involves Axon Guidance Genes. *PLOS Genetics* **12**, e1005950–e1005950 (2016).

80. Á. Raya, *et al.*, Notch activity acts as a sensor for extracellular calcium during vertebrate left–right determination. *Nature* **427**, 121–128 (2004).

81. J. Kimble, D. Hirsh, The postembryonic cell lineages of the hermaphrodite and male gonads in Caenorhabditis elegans. *Developmental Biology* **70**, 396–417 (1979).

82. M. C. Abraham, Y. Lu, S. Shaham, A Morphologically Conserved Nonapoptotic Program Promotes Linker Cell Death in Caenorhabditis elegans. *Developmental Cell* **12**, 73–86 (2007).

83. J. W. Ziel, D. R. Sherwood, Roles for netrin signaling outside of axon guidance: a view from the worm. *Developmental dynamics : an official publication of the American Association of Anatomists* **239**, 1296–305 (2010).

84. D. C. Bergmann, *et al.*, Embryonic handedness choice in C. elegans involves the Gα protein GPA-16. *Development* **130**, 5731–5740 (2003).

85. W. B. Wood, D. Bergmann, A. Florance, Maternal effect of low temperature on handedness determination in C-elegans embryos. *Developmental Genetics* **19**, 222–230 (1996).

86. D. C. Callander, M. R. Alcorn, B. Birsoy, J. H. Rothman, Natural reversal of left-right gut/gonad asymmetry in C. elegans males is independent of embryonic chirality. *Genesis* **52** (2014).

87. M. R. Alcorn, *et al.*, Heterotaxy in Caenorhabditis: Widespread natural variation in left - Right arrangement of the major organs. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371** (2016).

88. S. S. Chan, *et al.*, UNC-40 , a C . elegans Homolog of DCC (Deleted in Colorectal Cancer), Is Required in Motile Cells Responding to UNC-6 Netrin Cues. **87**, 187–195 (1996).

89. R. Y. Yu, C. Q. Nguyen, D. H. Hall, K. L. Chow, Expression of ram-5 in the structural cell is required for sensory ray morphogenesis in Caenorhabditis elegans male tail. *The EMBO journal* **19**, 3542–55 (2000).

90. J. E. Sulston, D. G. Albertson, J. N. Thomson, The Caenorhabditis elegans male: Postembryonic development of nongonadal structures. *Developmental Biology* **78**, 542–576 (1980).

91. Y. Yamaguchi, M. Miura, Programmed Cell Death in Neurodevelopment. *Developmental Cell* **32**, 478–490 (2015).

92. L. Lin, E. H. Baehrecke, Molecular & Cellular Oncology Autophagy, cell death, and cancer Autophagy, cell death, and cancer (2015) https:/doi.org/10.4161/23723556.2014.985913.

93. R. W. Oppenheim, Cell Death During Development of the Nervous System. *Annual Review of Neuroscience* **14**, 453–501 (1991).

94. R. I. Richards, *et al.*, The enemy within: Innate surveillance-mediated cell death, the common mechanism of neurodegenerative disease. *Frontiers in Neuroscience* **10**, 1–20 (2016).

95. R. R. Buss, W. Sun, R. W. Oppenheim, ADAPTIVE ROLES OF PROGRAMMED CELL DEATH DURING NERVOUS SYSTEM DEVELOPMENT. *Annual Review of Neuroscience* **29**, 1–35 (2006).

96. C. C. Chiu, *et al.*, Neuroinflammation in animal models of traumatic brain injury. *Journal of Neuroscience Methods* (2016) https:/doi.org/10.1016/j.jneumeth.2016.06.018.

97. H. Shojo, *et al.*, Genetic and histologic evidence implicates role of inflammation in traumatic brain injury-induced apoptosis in the rat cerebral cortex following moderate fluid percussion injury. *Neuroscience* (2010) https:/doi.org/10.1016/j.neuroscience.2010.10.018.

98. R. Zhao, *et al.*, Novel roles of apoptotic caspases in tumor repopulation, epigenetic reprogramming, carcinogenesis, and beyond. *Cancer Metastasis Rev* **37**, 227–236 (2018).

99. Z. Dezhina, S. Ranlund, M. Kyriakopoulos, S. C. R. Williams, D. Dima, A systematic review of associations between functional MRI activity and polygenic risk for schizophrenia and bipolar disorder. *Brain Imaging and Behavior* **13**, 862–877 (2019).

100. M. Solmi, *et al.*, Systematic review and meta-analysis of the efficacy and safety of minocycline in schizophrenia. *CNS Spectrums* **22**, 415–426 (2017).

101. H. M. Ellis, H. R. Horvitz, Genetic control of programmed cell death in the nematode C. elegans. *Cell* **44**, 817–829 (1986).

102. T. A. Jarrell, *et al.*, The Connectome of a Decision Making Neural Network. *Science* **in press**, 437–444 (2012).

103. G. Dalpe, H. Zheng, L. Brown, J. Culotti, Semaphorin-1 and netrin signal in parallel and permissively to position the male ray 1 sensillum in Caenorhabditis elegans. *Genetics* **192**, 959–971 (2012).

104. F. Nakao, *et al.*, The plexin PLX-2 and the ephrin EFN-4 have distinct roles in MAB-20/semaphorin 2A signaling in Caenorhabditis elegans morphogenesis. *Genetics* **176**, 1591–1607 (2007).

105. S. E. Baird, S. W. Emmons, Properties of a class of genes required for ray morphogenesis in Caenorhabditis elegans. *Genetics* **126**, 335–344 (1990).

106. S. E. Baird, D. H. Fitch, I. A. Kassem, S. W. Emmons, Pattern formation in the nematode epidermis: determination of the arrangement of peripheral sense organs in the C. elegans male tail. *Development* **113** (1991).

107. L. Jia, S. W. Emmons, Genes that control ray sensory neuron axon development in the Caenorhabditis elegans male. *Genetics* **173**, 1241–1258 (2006).

108. M. E. Sutherlin, S. W. Emmons, Selective lineage specification by mab-19 during Caenorhabditis elegans male peripheral sense organ development. *Genetics* **138**, 675–688 (1994).

109. M. G. Sterken, L. B. Snoek, J. E. Kammenga, E. C. Andersen, The laboratory domestication of Caenorhabditis elegans. *Trends in Genetics* **31**, 224–231 (2015).

110. M. Grosmaire, *et al.*, Males as somatic investment in a parthenogenetic nematode. *Science* **363**, 1210–1213 (2019).

111. K. Kiontke, D. H. a Fitch, The phylogenetic relationships of Caenorhabditis and other rhabditids. *WormBook : the online review of C. elegans biology*, 1–11 (2005).

112. K. C. Kiontke, *et al.*, A phylogeny and molecular barcodes for Caenorhabditis, with numerous new species from rotting fruits. *BMC Evolutionary Biology* **11**, 339–339 (2011).

113. G. Huelsz-Prince, J. S. van Zon, Canalization of C. elegans Vulva Induction against Anatomical Variability. *Cell Systems* **4**, 219-230.e6 (2017).

114. , Diversification of C. elegans Motor Neuron Identity via Selective Effector Gene Repression https:/doi.org/10.1016/j.neuron.2016.11.036.

115. O. Hobert, Architecture of a MicroRNA-controlled gene regulatory network that diversifies neuronal cell fates. *Cold Spring Harbor Symposia on Quantitative Biology* **71**, 181–188 (2006).

116. M. Doitsidou, *et al.*, A combinatorial regulatory signature controls terminal differentiation of the dopaminergic nervous system in C. elegans. *Genes and Development* **27**, 1391–1405 (2013).

117. J. Liang, *et al.*, The Caenorhabditis elegans schnurri homolog sma-9 mediates stage- and cell type-specific responses to DBL-1 BMP-related signaling. *Development (Cambridge, England)* **130**, 6453–64 (2003).

118. J. C. Downes, B. Birsoy, K. C. Chipman, J. H. Rothman, Handedness of a motor program in C. elegans is independent of left-right body asymmetry. *PloS one* **7**, e52138–e52138 (2012).

119. M. Hirnstein, M. Hausmann, O. Güntürkün, The evolutionary origins of functional cerebral asymmetries in humans: Does lateralization enhance parallel processing? *Behavioural Brain Research* **187**, 297–303 (2008).

120. W. C. McGrew, L. F. Marchant, Laterality of hand use pays off in foraging success for wild chimpanzees. *Primates* **40**, 509–513 (1999).

121. K. E. Wagner, L. M. Hopper, S. R. Ross, Asymmetries in the production of self-directed behavior by chimpanzees and gorillas during a computerized cognitive test. *Animal Cognition* **19**, 343–350 (2016).

122. A. K. Schnell, R. T. Hanlon, A. Benkada, C. Jozet-Alves, Lateralization of Eye Use in Cuttlefish: Opposite Direction for Anti-Predatory and Predatory Behaviors. *Frontiers in Physiology* **7**, 620–620 (2016).

123. L. J. Rogers, Evolution of hemispheric specialization: advantages and disadvantages. *Brain and language* **73**, 236–53 (2000).

124. O. Güntürkün, *et al.*, Asymmetry pays: Visual lateralization improves discrimination success in pigeons. *Current Biology* **10**, 1079–1081 (2000).

125. S. Brenner, The genetics of Caenorhabditis elegans. *Genetics* **77**, 71–94 (1974).

126. L. C. Lyons, R. M. Hecht, Acute Ethanol Exposure Induces Nondisjunction of the X Chromosome During Spermatogenesis. *Worm Breeder's Gazette* **14**, 52–52 (1997).

127. J. A. Lewis, C.-H. Wu, H. Berg, J. H. Levine, The Genetics of Levamisole Resistance in the Nematode CAENORHABDITIS ELEGANS. *Genetics* **95**, 905–905 (1980).

128. Y. Benjamini, Y. Hochberg, On the adaptive control of the false discovery fate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics* **25**, 60–83 (2000).

129. R. B. Canty A, boot: Bootstrap R (S-Plus) functions. R package version 1.3-17. (2015).

130. S. T. Buckland, A. C. Davison, D. V. Hinkley, Bootstrap Methods and Their Application. *Biometrics* **54**, 795–795 (1998).

131. R. Pruim, D. T. Kaplan, N. J. Horton, The mosaic package: Helping students to “think with data” using R. *R Journal* **9**, 77–102 (2017).

132. C. Ginestet, ggplot2: Elegant Graphics for Data Analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **174**, 245–246 (2011).

133. C. O. Wilke, *Streamlined Plot Theme and Plot Annotations for “ggplot2”* (2019).

134. N. Xiao, ggsci: Scientific Journal and Sci-Fi Themed Color Palettes for "ggplot2. *R package version 2.7* (2017).

135. C. M. Pfeffer, A. T. K. Singh, Apoptosis: A Target for Anticancer Therapy. *International Journal of Molecular Sciences* **19**, 448 (2018).

136. S. J. Martin, C. M. Henry, Distinguishing between apoptosis, necrosis, necroptosis and other cell death modalities. *Methods (San Diego, Calif.)* **61**, 87–9 (2013).

137. A. M. Chinnaiyan, D. Chaudhary, K. O’Rourke, E. V. Koonin, V. M. Dixit, Role of CED-4 in the activation of CED-3. *Nature* **388**, 728–729 (1997).

138. Q. A. Liu, M. O. Hengartner, Candidate adaptor protein CED-6 promotes the engulfment of apoptotic cells in C. elegans. *Cell* (1998) https:/doi.org/10.1016/S0092-8674(00)81202-7.

139. W. D. Fairlie, *et al.*, CED-4 forms a 2 : 2 heterotetrameric complex with CED-9 until specifically displaced by EGL-1 or CED-13. *Cell Death & Differentiation* **13**, 426–434 (2006).

140. J. Li, W. Le, Modeling neurodegenerative diseases in Caenorhabditis elegans. *Experimental Neurology* **250**, 94–103 (2013).

141. D. W. Bailey, Recombinant-inbred strains: An Aid to Finding Identity, Linkage, and Function of Histocompatibility and Other Genes. *Transplantation* **11**, 325–327 (1971).

142. K. W. Broman, A brief tour of R / qtl. 889–890 (2012).

143. H. M. Kang, *et al.*, Efficient Control of Population Structure in Model Organism Association Mapping. *Genetics* **178**, 1709–1723 (2008).

144. E. C. Andersen, *et al.*, Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic diversity. *Nature Genetics* **44**, 285–290 (2012).

145. E. C. Andersen, Supplemental Material. *Nature genetics* (2012) https:/doi.org/10.1038/ng.1050.

146. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

147. S. E. Baird, C. R. Davidson, J. C. Bohrer, The genetics of ray pattern variation in Caenorhabditis briggsae. *BMC Evolutionary Biology* **5** (2005).

148. P. Donnelly, Progress and challenges in genome-wide association studies in humans. *Nature* **456**, 728–31 (2008).

149. A. B. Paaby, M. V. Rockman, Cryptic genetic variation: evolution's hidden substrate. *Nature Reviews Genetics* **15**, 247–258 (2014).

150. R. Lints, L. Jia, K. Kim, C. Li, S. W. Emmons, Axial patterning of C. elegans male sensilla identities by selector genes. *Developmental Biology* **269**, 137–151 (2004).

151. M. Porta-de-la-Riva, L. Fontrodona, A. Villanueva, J. Cerón, Basic Caenorhabditis elegans methods: synchronization and observation. *Journal of visualized experiments : JoVE*, e4019–e4019 (2012).

152. A. C. Billi, S. E. J. Fischer, J. K. Kim, Endogenous RNAi pathways in C. elegans. *WormBook : the online review of C. elegans biology*, 1–49 (2014).

153. H. M. Kang, *et al.*, *emma: Efficient control of population structure in model organism association mapping (Efficient Mixed Model Association)*.

154. O. Hobert, The impact of whole genome sequencing on model system genetics: Get ready for the ride. *Genetics* **184**, 317–319 (2010).

155. Y. N. Torres Cleuren, *et al.*, Extensive intraspecies cryptic variation in an ancient embryonic gene regulatory network. *eLife* **8** (2019).

156. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

157. P. Danecek, *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

158. P. J. Bradbury, *et al.*, TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).

159. E. K. O. Namigai, N. J. Kenny, S. M. Shimeld, Right Across the Tree of Life : The Evolution of Left – Right Asymmetry in the Bilateria. **470**, 458–470 (2014).

160. R. Palmer, From symmetry to asymmetry: phylogenetic patterns of asymmetry variation in animals and their evolutionary significance. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 14279–14286 (1996).

161. J. K. Takeuchi, *et al.*, Baf60c is a nuclear Notch signaling component required for the establishment of left-right asymmetry. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 846–51 (2007).

162. J. L. Lohr, M. C. Danos, J. H. Yost, Left-right asymmetry of a nodal-related gene is regulated by dorsoanterior midline structures during Xenopus development. *Development* **124**, 1465–1472 (1997).

163. M. Levin, *et al.*, Left/right patterning signals and the independent regulation of different aspects of situs in the chick embryo. *Developmental biology* **189**, 57–67 (1997).

164. M. Logan, S. M. Pagán-Westphal, D. M. Smith, L. Paganessi, C. J. Tabin, The transcription factor Pitx2 mediates situs-specific morphogenesis in response to left-right asymmetric signals. *Cell* **94**, 307–17 (1998).

165. R. Kuroda, B. Endo, M. Abe, M. Shimizu, Chiral blastomere arrangement dictates zygotic left–right asymmetry pathway in snails. *Nature* **462**, 790–794 (2009).

166. S. Nakano, B. Stillman, H. R. Horvitz, Replication-coupled chromatin assembly generates a neuronal bilateral asymmetry in C. elegans. *Cell* **147**, 1525–36 (2011).

167. S. Schonegg, A. A. Hyman, W. B. Wood, Timing and mechanism of the initial cue establishing handed left-right asymmetry in Caenorhabditis elegans embryos. *Genesis* **52**, 572–580 (2014).

168. A. F. Ramsdell, Left-right asymmetry and congenital cardiac defects: Getting to the heart of the matter in vertebrate left-right axis determination. *Developmental Biology* **288**, 1–20 (2005).

169. C. B. Arrington, A. G. Peterson, H. J. Yost, Sdc2 and Tbx16 regulate Fgf2-dependent epithelial cell morphogenesis in the ciliated organ of asymmetry. *Development (Cambridge, England)* **140**, 4102–9 (2013).

170. R. Lagadec, *et al.*, The ancestral role of nodal signalling in breaking L/R symmetry in the vertebrate forebrain. *Nature communications* **6**, 6686–6686 (2015).

171. J. M. Neugebauer, H. Joseph Yost, FGF signaling is required for brain left-right asymmetry and brain midline formation. *Developmental Biology* **386**, 123–134 (2014).

172. O. Hobert, Development of left/right asymmetry in the Caenorhabditis elegans nervous system: From zygote to postmitotic neuron. *Genesis* **52**, 528–543 (2014).

173. C. Francks, Exploring human brain lateralization with molecular genetics and genomics. *Annals of the New York Academy of Sciences* **1359**, 1–13 (2015).

174. L. J. Rogers, Asymmetry of brain and behavior in animals: Its development, function, and human relevance. *Genesis* **52**, 555–571 (2014).

175. H. J. Yost, Left-right development from embryos to brains. *Developmental Genetics* **23**, 159–163 (1998).

176. S. J. Kim, Heterotaxy syndrome. *Korean Circulation Journal* **41**, 227–232 (2011).

177. M. W. Leigh, *et al.*, Clinical and genetic aspects of primary ciliary dyskinesia/Kartagener syndrome. *Genetics in medicine : official journal of the American College of Medical Genetics* **11**, 473–87 (2009).

178. G. J. Hermann, B. Leung, J. R. Priess, Left-right asymmetry in C. elegans intestine organogenesis involves a LIN-12/Notch signaling pathway. *Development (Cambridge, England)* **127**, 3429–3440 (2000).

179. C. Pohl, Z. Bao, Chiral Forces Organize Left-Right Patterning in C. elegans by Uncoupling Midline and Anteroposterior Axis. *Developmental Cell* **19**, 402–412 (2010).

180. Z. Bao, Z. Zhao, T. J. Boyle, J. I. Murray, R. H. Waterston, Control of cell cycle timing during C. elegans embryogenesis. *Developmental Biology* **318**, 65–72 (2008).

181. R. J. Poole, O. Hobert, Early Embryonic Programming of Neuronal Left/Right Asymmetry in C. elegans. *Current Biology* **16**, 2279–2292 (2006).

182. P. J. Heid, *et al.*, The zinc finger protein DIE-1 is required for late events during epithelial cell rearrangement in C. elegans. *Developmental biology* **236**, 165–80 (2001).

183. O. Hobert, R. J. Johnston, S. Chang, Left–right asymmetry in the nervous system: the Caenorhabditis elegans model. *Nature Reviews Neuroscience* **3**, 629–640 (2002).

184. M. V. Rockman, L. Kruglyak, Breeding designs for recombinant inbred advanced intercross lines. *Genetics* **179**, 1069–78 (2008).

185. M. V. Rockman, L. Kruglyak, Recombinational landscape and population genomics of Caenorhabditis elegans. *PLoS genetics* **5**, e1000419–e1000419 (2009).

186. F. Duveau, M.-A. Félix, Role of Pleiotropy in the Evolution of a Cryptic Developmental Variation in Caenorhabditis elegans. *PLoS Biology* **10**, e1001230–e1001230 (2012).

187. J. Milloz, F. Duveau, I. Nuez, M.-A. Felix, Intraspecific evolution of the intercellular signaling network underlying a robust developmental system. *Genes & Development* **22**, 3064–3075 (2008).

188. C. J. Malone, W. D. Fixsen, H. R. Horvitz, M. Han, UNC-84 localizes to the nuclear envelope and is required for nuclear migration and anchoring during C. elegans development. *Development (Cambridge, England)* **126**, 3171–3181 (1999).

189. J. M. VanLiere, N. a Rosenberg, Mathematical properties of the measure of linkage disequilibrium. *Theoretical Population Biology* **74**, 130–137 (2008).

190. A. Charmantier, D. Garant, Environmental quality and evolutionary potential: lessons from wild populations. *Proceedings of the Royal Society B: Biological Sciences* **272**, 1415–1425 (2005).

191. D. L. Byers, Components of phenotypic variance. *Nature Education* **1**, 1–5 (2008).

192. N. Wray, P. Visscher, Estimating trait heritability. *Nature Education* **1**, 29–29 (2008).

193. M. V. Rockman, S. S. Skrovanek, L. Kruglyak, Selection at linked sites shapes heritable phenotypic variation in C. elegans. *Science (New York, N.Y.)* **330**, 372–376 (2010).

194. M. Soller, J. S. Beckmann, Marker-based mapping of quantitative trait loci using replicated progenies. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **80**, 205–8 (1990).

195. M. Akerfelt, R. I. Morimoto, L. Sistonen, Heat shock factors: integrators of cell stress, development and lifespan. *Nature reviews. Molecular cell biology* **11**, 545–55 (2010).

196. L. N. Vandenberg, M. Levin, Far From Solved : A Perspective on What We Know About Early Mechanisms of Left – Right Asymmetry. 3131–3146 (2010).

197. J. R. Priess, J. N. Thomson, Cellular interactions in early C. elegans embryos. *Cell* **48**, 241–250 (1987).

198. A. Davison, *et al.*, Formin Is Associated with Left-Right Asymmetry in the Pond Snail and the Frog. *Current Biology* **26**, 654–660 (2016).

199. C. Pohl, Left-right patterning in the C. elegans embryo: Unique mechanisms and common principles. *Communicative and Integrative Biology* **4**, 1–7 (2011).

200. C.-P. Heisenberg, Y. Bellaïche, Forces in tissue morphogenesis and patterning. *Cell* **153**, 948–62 (2013).

201. G. Calenda, *et al.*, Identification of two new members, XPLAC and XTES, of the XK family. *Gene* **370**, 6–16 (2006).

202. D. D. Shaye, I. Greenwald, The disease-associated formin INF2/EXC-6 organizes lumen and cell outgrowth during tubulogenesis by regulating F-actin and microtubule cytoskeletons. *Developmental Cell* **32**, 743–755 (2015).

203. M. Buechner, D. H. Hall, H. Bhatt, E. M. Hedgecock, Cystic canal mutants in Caenorhabditis elegans are defective in the apical membrane domain of the renal (excretory) cell. *Developmental biology* **214**, 227–41 (1999).

204. K. a Swan, *et al.*, cyk-1: a C. elegans FH gene required for a late step in embryonic cytokinesis. *Journal of cell science* **111 (Pt 1**, 2017–2027 (1998).

205. A. F. Severson, D. L. Baillie, B. Bowerman, A Formin Homology protein and a profilin are required for cytokinesis and Arp2/3-independent assembly of cortical microfilaments in C. elegans. *Current Biology* **12**, 2066–2075 (2002).

206. K. E. Watkins, *et al.*, Structural Asymmetries in the Human Brain: a Voxel-based Statistical Analysis of 142 MRI Scans. *Cerebral Cortex* **11**, 868–877 (2001).

207. L. Cochella, *et al.*, Two distinct types of neuronal asymmetries are controlled by the Caenorhabditis elegans zinc finger transcription factor die-1. *Genes and Development* **28**, 34–43 (2014).

208. E. R. Troemel, A. Sagasti, C. I. Bargmann, Lateral signaling mediated by axon contact and calcium entry regulates asymmetric odorant receptor expression in C. elegans. *Cell* **99**, 387–98 (1999).

209. R Core Team, *R: A language and environment for statistical computing* (R Foundation for Statistical Computing, 2019).

210. S. Purcell, *et al.*, PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559–575 (2007).

211. A. Sagasti, *et al.*, The CaMKII UNC-43 Activates the MAPKKK NSY-1 to Execute a Lateral Signaling Decision Required for Asymmetric Olfactory Neuron Fates. *Cell* **105**, 221–232 (2001).

212. C. F. Chuang, C. I. Bargmann, A Toll-interleukin 1 repeat protein at the synapse specifies asymmetric odorant receptor expression via ASK1 MAPKKK signaling. *Genes and Development* **19**, 270–281 (2005).

213. K. C. Reddy, E. C. Andersen, L. Kruglyak, D. H. Kim, A polymorphism in npr-1 is a behavioral determinant of pathogen susceptibility in C. elegans. *Science (New York, N.Y.)* **323**, 382–384 (2009).

214. L. M. Noble, *et al.*, Natural Variation in plep-1 Causes Male-Male Copulatory Behavior in C. elegans. *Current Biology* **25**, 2730–2737 (2015).

215. L. B. Snoek, *et al.*, Widespread Genomic Incompatibilities in Caenorhabditis elegans. *G3: Genes*|*Genomes*|*Genetics* **4**, 1813–1823 (2014).

216. K. L. Howe, B. J. Bolt, M. Shafie, P. Kersey, M. Berriman, WormBase ParaSite − a comprehensive resource for helminth genomics. *Molecular and Biochemical Parasitology* **215**, 2–10 (2017).

217. C. Chen, L. a Fenk, M. de Bono, Efficient genome editing in Caenorhabditis elegans by CRISPR-targeted homologous recombination. *Nucleic acids research* **41**, e193–e193 (2013).

218. M. Gendrel, E. G. Atlas, O. Hobert, A cellular and regulatory map of the GABAergic nervous system of C. elegans. *Elife* **5** (2016).

219. H. Zou, T. Hastie, Regularization and variable selection via the elastic-net. *Journal of the Royal Statistical Society* **67**, 301–320 (2005).

220. H. Zou, T. Hastie, Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 301–320 (2017).

221. T. Hastie, R. Tibshirani, J. Friedman, *Elements of Statistical Learning: data mining, inference, and prediction. 2nd Edition.* (January 28, 2020).

222. V. Anttila, *et al.*, Genome-wide meta-analysis identifies new susceptibility loci for migraine. *Nature genetics* **45**, 912–7 (2013).

223. S. H. Lee, *et al.*, Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature genetics* **45**, 984–94 (2013).

224. Z. Zhao, *et al.*, Transcriptome sequencing and genome-wide association analyses reveal lysosomal function and actin cytoskeleton remodeling in schizophrenia and bipolar disorder. *Molecular Psychiatry* **20**, 563–572 (2015).

225. N. Craddock, *et al.*, Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* **464**, 713–20 (2010).

226. S. Ueno, J. Kuo, *Advances in the Understanding of Biological Sciences Using Next Generation Sequencing (NGS) Approaches* (2014) https:/doi.org/10.1007/978-3-319-17157-9.

227. D. Hamer, L. Sirota, Beware the chopsticks gene. *Molecular Psychiatry* **5**, 11–13 (2000).

228. K. Miclaus, *et al.*, Variability in GWAS analysis: the impact of genotype calling algorithm inconsistencies. *The pharmacogenomics journal* **10**, 324–35 (2010).

229. C. Papachristou, C. Ober, M. Abney, A LASSO penalized regression approach for genome-wide association analyses using related individuals : application to the Genetic Analysis Workshop 19 simulated data. *BMC Proceedings* **10**, 0–5 (2016).

230. P. Waldmann, G. Mészáros, B. Gredler, C. Fuerst, J. Sölkner, Evaluation of the lasso and the elastic net in genome-wide association studies. **4**, 1–11 (2013).

231. E. Cule, P. Vineis, M. De Iorio, Significance testing in ridge regression for genetic data. *BMC Bioinformatics* **12**, 372–372 (2011).

232. T. Fearn, Ridge regression. **24**, 18–18 (2013).

233. J. O. Ogutu, T. Schulz-Streeck, H.-P. Piepho, Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. *BMC Proceedings* **6**, S10–S10 (2012).

234. B. Li, *et al.*, Genomic Prediction of Breeding Values Using a Subset of SNPs Identified by Three Machine Learning Methods. *Front Genet* **9**, 237 (2018).

235. F. Pedregosa, *et al.*, Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825−2830 (2011).

236. J. Zou, *et al.*, A primer on deep learning in genomics. *Nat Genet* **51**, 12–18 (2019).

237. R. Tibshirani, Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **58**, 267–288 (1996).

238. L. Byerly, R. C. Cassada, R. L. Russell, Life cycle of the nematode C. elegans. *Developmental Biology* **51**, 23–33 (1976).

239. R. S. Kamath, *et al.*, Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature* **421**, 231–237 (2003).

240. D. J. Dickinson, B. Goldstein, CRISPR-based methods for caenorhabditis elegans genome engineering. *Genetics* **202**, 885–901 (2016).

241. M. F. Maduro, Gut development in C. elegans. *Seminars in Cell & Developmental Biology* **66**, 3–11 (2017).

242. M. W. Libbrecht, W. S. Noble, Machine learning applications in genetics and genomics. *Nature Reviews Genetics* **16**, 321–332 (2015).

243. P. L. Tan, S. C. Tan, C. P. Lim, S. E. Khor, A Modified Two-Stage SVM-RFE Model for Cancer Classification Using Microarray Data Gene Selection Algorithm and Support Vector Machine. 668–675 (2011).

244. Z. Y. Algamal, M. H. Lee, Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification. *Computers in Biology and Medicine* **67**, 136–145 (2015).

245. K. Y. Yip, C. Cheng, M. Gerstein, Machine learning and genome annotation: a match meant to be? *Genome biology* **14**, 205–205 (2013).